



HAL
open science

Etude de la régulation de l'expression des gènes chez Escherichia coli, en lien avec le transport et le métabolisme du glucose : comprenant : - l'analyse moléculaire de la région promoteur dirigeant l'expression des gènes gapA et yeaA, - l'étude de l'effet de l'absence de protéine EIIBCGlc assurant le transport spécifique du glucose sur le transcriptome d'Escherichia coli.

Benoit Thouvenot

► **To cite this version:**

Benoit Thouvenot. Etude de la régulation de l'expression des gènes chez Escherichia coli, en lien avec le transport et le métabolisme du glucose : comprenant : - l'analyse moléculaire de la région promoteur dirigeant l'expression des gènes gapA et yeaA, - l'étude de l'effet de l'absence de protéine EIIBCGlc assurant le transport spécifique du glucose sur le transcriptome d'Escherichia coli.. Sylviculture, foresterie. Université Henri Poincaré - Nancy 1, 2004. Français. NNT : 2004NAN10191 . tel-01748108

HAL Id: tel-01748108

<https://hal.univ-lorraine.fr/tel-01748108v1>

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



UNIVERSITÉ HENRI POINCARÉ NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54500 VILLERS-LÈS-NANCY

FACULTE DES SCIENCES & TECHNIQUES

U.F.R Sciences et Techniques Biologiques
Ecole Doctorale Biologie, Santé et environnement

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Biologie Moléculaire

par **Benoit THOUVENOT**

Etude de la régulation de l'expression des gènes chez *Escherichia coli*, en lien avec le transport et le métabolisme du glucose, comprenant :

- l'analyse moléculaire de la région promoteur dirigeant l'expression des gènes *gapA* et *yeaA*,
- l'étude de l'effet de l'absence de protéine EIIBC^{Glc} assurant le transport spécifique du glucose sur le transcriptome d'*Escherichia coli*.

Soutenance publique prévue le 25 Novembre 2004 devant la commission d'examen :

Membres du jury :

Président du jury :	Mr F. Barras	Professeur, Marseille
Rapporteurs :	Mr M. Buckle	Directeur de Recherche CNRS, Cachan
	Mme J. Plumbridge	Directeur de Recherche CNRS, Paris
Examineur :	Mr A. Visvikis	Professeur, Université Henri Poincaré, Nancy I
Directeur de thèse :	Mme C. Branlant	Directeur de recherche CNRS, Nancy

BIBLIOTHEQUE SCIENCES NANCY 1



Travail ré. D **095 168321 5** ...s ARN et Enzymologie Moléculaire UMR CNRS-UHP 7567

Faculté des Sciences & Techniques - BP 236 - 54506 Vandoeuvre-lès-Nancy Cedex

Monsieur **THOUVENOT Benoît**

DOCTORAT de l'UNIVERSITE HENRI POINCARÉ, NANCY 1
en BIOLOGIE MOLECULAIRE

VU, APPROUVÉ ET PERMIS D'IMPRIMER ၅၈၅

Nancy, le 8 décembre 2004

Le Président de l'Université



Je tiens à exprimer ma profonde gratitude à Madame Christiane Branlant pour m'avoir accueilli dans son laboratoire, pour avoir dirigé ce travail et pour m'avoir fait découvrir le monde de la recherche. Je la remercie pour la formation scientifique qu'elle a su me donner. La rédaction de ce mémoire doit beaucoup à ses conseils, sa rigueur scientifique et les discussions que nous avons eues et le temps très important qu'elle y a consacré.

Je suis très reconnaissant à Madame Jacqueline Plumbidge et Monsieur Malcolm Buckle d'avoir accepté d'être les rapporteurs de cette thèse. Je tiens aussi à remercier Monsieur Malcolm Buckle pour m'avoir accueilli deux fois pour des séjours dans son équipe et m'avoir montré la puissance de la technique laser pour l'étude des interactions ADN-ARN polymérase. Je le remercie tout particulièrement pour les discussions scientifiques que nous avons eues au cours de mon stage à Cachan. Je voudrais aussi remercier Bianca Sclavi pour son aide, ses conseils et les discussions que nous avons eues au cours de mes stages à Cachan mais aussi ensuite à mon retour à Nancy.

Je tiens à remercier Monsieur Frédéric Barras de m'avoir fait l'honneur d'accepter d'être membre du jury de ma thèse et pour m'avoir communiqué des résultats non publiés de son équipe. Je tiens également à remercier Monsieur Athanase Visvikis pour avoir accepté de faire partie de mon jury de thèse.

Un grand merci à Bruno Charpentier pour m'avoir encadré au cours de ma thèse. Ses précieux conseils et ses idées m'ont permis d'avancer tout au long de ce travail. Je le remercie aussi pour tout ce qu'il a fait pour moi en dehors du laboratoire.

Je tiens aussi à remercier le Pr. G. Branlant qui m'a permis d'avoir un rôle actif dans l'enseignement à l'UHP d'abord en tant que moniteur puis en temps qu'ATER.

Je remercie également mes collègues qui ont travaillé avec moi sur la thématique E. coli au laboratoire, et en particulier Nathalie Robas, Vincent Bardey et Corinne Vallét pour notre collaboration et nos nombreuses discussions scientifiques durant notre période commune de rédaction. Je remercie aussi les autres collègues du laboratoire pour l'ambiance qu'ils ont su créer et en particulier Antoine, Jean-Michel, Alan, Xavier et Sebastien mais aussi Delphine R, Delphine M, les « enzymo » Olivier, Alex, François...

Je voudrais enfin remercier mes amis, ma famille et tout particulièrement mes parents pour leur soutien inconditionnel et leurs conseils tout au long de mes études. Que serait ma vie sans ma femme Emmanuelle à qui je dois beaucoup, surtout le petit bébé qui va venir au monde et illuminer notre vie.

Abbréviations

A	: adénine
ADN	: acide désoxyribonucléique
ADNc	: acide désoxyribonucléique complémentaire
ADP	: adénosine 5' diphosphate
AMP	: adénosine 5' monophosphate
AMPc	: adénosine monophosphate cyclique
ARI/II	: "Activating Region" I/II
ARN	: acide ribonucléique
ARNm	: acide ribonucléique messenger
ARNnc	: ARN non codant
ARN pol	: ARN polymérase
ARNr	: acide ribonucléique ribosomique
ATP	: adénosine 5' triphosphate
C	: cytosine
Ci	: curie
cpm	: coups par minute
CTP	: cytidine 5' triphosphate
Da/kDa	: dalton/kilodalton
ddNTP	: didésoxyribonucléoside 5' triphosphate
DMS	: diméthyle sulfate
DNase	: désoxyribonucléase
dNTP	: désoxyribonucléoside triphosphate
DO	: densité optique
DR1/2	: Dispensable Region 1/2
DTT	: dithiothréitol
dUMP	: désoxyuridine 5' monophosphate
dUTP	: désoxyuridine 5' triphosphate
ECF	: Extra Cytoplasmic Function
EIIA ^{Glc}	: Domaine A de l'enzyme II du système PTS du glucose
EIIBC ^{Glc}	: Domaines B et C de l'enzyme II du système PTS du glucose
E4PDH	: érythrose 4-phosphate déshydrogénase
EDTA	: acide éthylène diamine tétra-acétique
EtOH	: éthanol
E σ^{24}	: ARN polymérase holoenzyme ayant le facteur σ^{24}
E σ^{28}	: ARN polymérase holoenzyme ayant le facteur σ^{28}
E σ^{32}	: ARN polymérase holoenzyme ayant le facteur σ^{32}
E σ^{38}	: ARN polymérase holoenzyme ayant le facteur σ^{38}
E σ^{54}	: ARN polymérase holoenzyme ayant le facteur σ^{54}
E σ^{70}	: ARN polymérase holoenzyme ayant le facteur σ^{70}
FBA	: Fructose 1,6-biphosphate aldolase
G3P	: glycéraldéhyde 3-phosphate
g/mg/ μ g/ng	: gramme/milligramme/microgramme/nanogramme
GAPDH	: glycéraldéhyde 3-phosphate déshydrogénase
GDP	: guanosine 5' diphosphate
GTP	: guanosine 5' triphosphate
HK	: Histidine Kinase
K _B	: constante thermodynamique de formation du complexe fermé d'initiation de la transcription
kb	: kilobase

$k_1/k_2/k_3$: constantes cinétiques de formation des complexes d'initiation de la transcription
l/ml/ μ l	: litre/millilitre/microlitre
LB/LBA	: Luria Bertani/Luria Bertani supplémenté en ampicilline
M/mM/ μ M	: molaire/millimolaire/micromolaire
mA	: milliampère
MFD	: Mutation Frequency Decline
min	: minute
mol/mmol/ μ mol	: mole/millimole/micromole
N	: nucléotide
NAD	: nicotinamide adénine dinucléotide
NADP	: nicotinamide adénine dinucléotide phosphate
NCD1/2	: Non Conserved Domain 1/2
nm	: nanomètre
NTP	: ribonucléosides 5' triphosphate
nt	: nucléotides
ONPG	: O-Nitrophényl β -D-galactopyranoside
ORF	: "open reading frame"
pb	: paire de bases
pBS	: phagemide bluescript
p/v	: poids/volume
PEG	: polyéthylène glycol
PEP	: phospho-énol pyruvate
PGK	: phosphoglycérate kinase
PM	: poids moléculaire
PTS	: système phosphotransférase PEP dépendant
RBS	: "ribosome binding site"
RNase	: ribonucléase
RPc	: complexe fermé d'initiation de la transcription
RPélong	: complexe d'élongation de la transcription
RPint	: complexe intermédiaire d'initiation la transcription
RPinit	: complexe d'initiation de la transcription
RPo	: complexe ouvert d'initiation de la transcription
RR	: Régulateur de réponse
SDS	: sodium dodécylsulfate
TBE	: Tris-borate-EDTA
TE	: Tris-EDTA
TIR	: "Translation Initiation Region"
Tn	: transposon
tr/min	: tours/minute
Tris	: Tris (hydroxyméthyl) aminométhane
U	: unité enzymatique
UP	: "upstream region"
UTP	: uridine 5' triphosphate
UTR	: "untranslated region"
UV	: ultra-violet
V	: volt
V/cm	: Volt/centimètre
v/v	: volume/volume
W	: watt
X-gal	: 5-bromo-4-chloro-3-indoxyl- β -D-galactopyranoside

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

Sommaire

Avant-propos

Introduction

Partie 1 : La transcription chez <i>Escherichia coli</i>	1
I.) <u>Le mécanisme de l'initiation de la transcription chez les bactéries</u>.....	2
I.1) Les acteurs	2
I.1.1) Les promoteurs.....	2
I.1.2) L'ARN polymérase bactérienne (EC 2.7.7.6).....	5
I.1.2.1) L'apoenzyme.....	5
I.1.2.2) L'holoenzyme.....	9
I.1.2.2.1) Les facteurs σ	9
I.1.2.2.1.1) Le facteur σ^{70} chez <i>E. coli</i>	11
I.1.2.2.1.2) Les autres facteurs σ de la famille σ^{70}	14
I.1.2.2.1.3) Le facteur σ^{54} chez <i>E. coli</i>	15
I.1.2.2.2) L'holoenzyme $E\sigma^{70}$	15
I.1.2.2.3) Les autres holoenzymes.....	18
I.2) Le mécanisme d'initiation de la transcription	18
I.2.1) Mécanisme d'action de l'holoenzyme $E\sigma^{70}$	19
I.2.1.1) Processus de formation des complexes RPo.....	21
I.2.1.2) Le complexe RPo.....	27
I.2.1.3) Processus de formation du complexe RPelong.....	29
I.2.1.4) Le complexe RPelong et mécanisme d'élongation.....	31
I.2.1.5) La terminaison de la transcription.....	32
I.2.2) Mécanisme pour l'holoenzyme $E\sigma^{54}$	33
II.) <u>Régulation de l'initiation de la transcription</u>.....	34
II.1) Contrôle de l'initiation de la transcription par le choix du facteur σ	35
II.1.1) Le facteur σ^S	35
II.1.2) Le facteur σ^{32}	37
II.1.3) Le facteur σ^{28}	37
II.1.4) Les facteurs $\sigma^{24(E)}$ et σ^{FecI}	38
II.1.5) Conséquences des variations de concentration des facteurs σ sur le taux des différentes holoenzymes dans la cellule.....	38
II.2) Implication d'éléments en <i>cis</i> dans la modulation de la force d'un promoteur reconnu par l'holoenzyme $E\sigma^{70}$	40
II.2.1) Homologie de la séquence des boîtes -10 et -35 avec le consensus.....	40
II.2.2) Influence des séquences autour des boîtes -10 et -35.....	41
II.2.2.1) Séquence en amont de la boîte -35.....	42
II.2.2.2) Séquence de la région interboîte.....	43
II.2.2.3) Séquences en aval de la boîte -10.....	44
II.2.3) Influence de la structure de l'ADN.....	45
II.2.3.1) La courbure de l'ADN.....	46
II.2.3.1.1) La formation des courbures au sein de l'ADN.....	46

II.2.3.1.2) Influence de la présence de courbure sur l'efficacité de l'initiation de la transcription.....	48
II.2.3.1.2.1) Présence d'une courbure entre les boîtes -10 et -35.	48
II.2.3.1.2.2) Présence d'une courbure en amont de la boîte -35.	50
II.2.3.2) Le surenroulement de l'ADN.....	52
II.2.3.2.1) Définition du surenroulement.	52
II.2.3.2.2) Effet du taux de surenroulement sur l'initiation de la transcription.....	53
II.3) Régulation par des facteurs agissant en <i>trans</i>.	55
II.3.1) Régulation par des molécules de petite taille.....	55
II.3.1.1) Concentration en nucléotide initiateur.	55
II.3.1.2) Concentration en ppGpp.	56
II.3.2) Régulation par un facteur protéique.....	58
II.3.2.1) Modulation de la fixation des facteurs agissant en <i>trans</i> sur les séquences régulatrices.	59
II.3.2.1.1) Variation de la concentration en facteurs de transcription.....	59
II.3.2.1.2) Modification par liaison d'un cofacteur : exemple du complexe CRP-AMPC.	60
II.3.2.1.3) Modulation de la capacité de fixation par phosphorylation.....	63
II.3.2.1.4) Modulation de la capacité de fixation par séquestration.	63
II.3.2.2) Régulation négative.	64
II.3.2.3) Régulations positives.	66
II.3.2.3.1) Les activateurs de classe I.....	66
II.3.2.3.2) Les activateurs de classe II.	68
II.3.2.4) Modèles complexes.....	69
II.3.2.4.1) Régulation reposant sur un taux d'occupation différentiel des séquences régulatrices.....	70
II.3.2.4.2) Régulation par modification de la structure de l'ADN.....	70
II.3.2.4.3) Intervention de plusieurs régulateurs et intégration des signaux.....	71

Partie 2 : Adaptation d'*E. coli* aux changements environnementaux.....

74

I.) <u>Les porines bactériennes</u>	75
I.1) Les porines majeures	75
I.2) Les porines spécifiques	76
II.) <u>Le système PTS chez <i>E. coli</i></u>	78
II.1) Organisation protéique et génétique des composants du système PTS	79
II.1.1) Organisation protéique.....	79
II.1.1.1) La protéine EI.	79
II.1.1.2) La protéine HPr.....	79
II.1.1.3) Les complexes EII spécifiques et le mécanisme	

de translocation.....	80
II.1.2) Organisation des gènes du système PTS.....	82
II.2) Régulations exercées par le système PTS.....	82
II.2.1) Régulations par les composants du système PTS.....	83
II.2.1.1) Régulation exercée par EI.....	83
II.2.1.2) Régulation exercée par EIIA ^{Glc}	83
II.2.1.2.1) Phénomène d'exclusion de l'inducteur.....	83
II.2.1.2.2) Modulation de l'activité de l'adénylate cyclase.....	86
II.2.1.3) Modulation de l'activité du répresseur Mlc par EIIBC ^{Glc}	88
II.2.2) Les régulations de l'expression des gènes du système PTS.....	90
II.2.2.1) Régulation de l'expression des gènes impliqués dans le catabolisme d'autres sucres PTS que le glucose.....	90
II.2.2.2) Régulation de la transcription du gène <i>ptsG</i> et de l'opéron <i>pts</i>	91
II.2.2.3) Régulation de la transcription des gènes <i>crp</i> et <i>cya</i>	95
III.) Les systèmes à deux composants.....	97
III.1) L'organisation génomique des systèmes à deux composants.....	97
III.1.1) Chez les bactéries.....	98
III.1.2) Chez <i>E. coli</i>	98
III.2) Structure et fonction des systèmes à deux composants d' <i>E. coli</i>.....	99
III.2.1) Le composant histidine kinase.....	99
III.2.2) Régulateurs de réponse.....	100
III.3) Exemples de mécanismes de transfert et de régulations.....	102
III.3.1) Système EnvZ et OmpR.....	102
III.3.1.1) Mécanisme permettant la régulation de la concentration de OmpR-P.....	102
III.3.1.2) Sites de fixation de OmpR-P sur l'ADN et régulation de la synthèse des porines.....	103
III.3.2) Le chimiotactisme chez <i>E. coli</i>	104
III.3.2.1) Les flagelles.....	105
III.3.2.2) Le système CheA/CheY.....	106
Partie 3 : Présentation du sujet.....	109
I.) <u>Le gène <i>gapA</i> d'<i>E. coli</i></u>	110
I.1) Présentation de l'enzyme : la GAPDH (EC 1.2.1.13).....	110
I.2) Transcription du gène <i>gapA</i> codant la GAPDH d' <i>E. coli</i>	111
II.) <u>L'opéron <i>epd-pgk-fbaA</i> d'<i>E. coli</i></u>	113
III.) <u>Objectifs du travail</u>	114

Résultats-Discussion

Partie 1 : Etude du promoteur <i>gapA</i> P1 d' <i>E. coli</i>	116
I.) <u>Etude de l'architecture du promoteur <i>gapA</i> P1 et du complexe ouvert (RPo) formé à ce promoteur.</u>	116
I.1) Intérêt du sujet et état des travaux lorsque nous avons débuté l'étude.	116
I.2) Les objectifs de mes travaux, la démarche expérimentale et les résultats marquants obtenus.	118
I.2.1) Démonstration du fait que le promoteur <i>gapA</i> P1 fait partie de la famille des promoteurs étendus.....	118
I.2.2) Démonstration du fait que le promoteur <i>gapA</i> P1 est un promoteur étendu particulier nécessitant la présence d'une boîte -35.....	119
I.2.3) Démonstration du fait que les interactions entre l'ARN pol et l'ADN du promoteur <i>gapA</i> P1 sont multiples.....	120
I.2.4) Démonstration du rôle d'une courbure naturelle induite par deux séries de résidus T présentes dans la région interboîte et la boîte -10.	122
I.2.5) Démonstration du fait que les mutations créées influencent le taux de complexe RPo.....	122
II.) <u>Article 1 : The strong efficiency of the <i>Escherichia coli gapA</i> P1 promoter depends on a complex combination of functional determinants.</u> B. Thouvenot, B. Charpentier et C. Branlant <i>Biochem. J.</i> , 383 (2), 371-382.	123
III.) <u>Travaux complémentaires réalisés sur le complexe RPo formé avec le promoteur <i>gapA</i> P1.</u>	124
III.1) Présentation de la technique de pontage covalent par irradiation laser.	124
III.2) Application de la technique de pontage covalent par irradiation laser à l'étude du promoteur <i>gapA</i> P1 sauvage.	125
III.3) Application de la technique de pontage covalent par irradiation laser à l'étude des variants du promoteur <i>gapA</i> P1.	128
III.4) Etude de la stabilité des complexes RPo formés au promoteur <i>gapA</i> P1.	129
III.5) L'efficacité du promoteur <i>gapA</i> P1 dépend du taux de surenroulement de l'ADN.	133
Partie 2 : Etude de l'ensemble de la région intergénique <i>gapA-yeaA</i>	135
I.) <u>Intérêt d'une étude de la région intergénique <i>gapA-yeaA</i>.</u>	135
II.) <u>Rôle des sites de fixation de la protéine Fis dans la régulation de la transcription des ORF <i>gapA</i> et <i>yeaA</i>.</u>	136
II.1) Etude de l'effet de la protéine Fis sur l'activité du promoteur <i>gapA</i> P1.	136
II.2) Etude de l'effet de la protéine Fis sur l'activité des promoteurs <i>yeaA</i> P1 et <i>yeaA</i> P2.	138

III.) <u>Le promoteur <i>yeaA</i> P1 fait partie de la famille des promoteurs étendus.</u>	141
IV.) <u>Rôle de la forte courbure et du site de fixation du complexe CRP-AMPC présents dans la région intergénique <i>gapA-yeaA</i> sur l'efficacité des promoteurs <i>yeaA</i> P1 et P2.</u>	143
V.) <u>Etude sur l'implication possible de l'ARN régulateur RyhB dans la régulation de la traduction du transcrit <i>yeaA</i>.</u>	146
V.1) Présentation de l'ARN antisens RyhB et de son implication dans l'homéostasie du fer.	147
V.2) Recherche de complémentarités possibles entre l'ARNnc RyhB et l'ARNm <i>yeaA</i>.	149
V.3) Effet de la surexpression de l'ARN RyhB sur l'expression du gène <i>yeaA</i>.	151
V.4) Test d'un effet possible de la protéine Hfq sur la formation d'une interaction entre l'ARN RyhB et l'ARNm <i>yeaA</i>.	152
V.4.1) Purification de la protéine recombinante Hfq et vérification de son activité de fixation à l'ARN RyhB.	152
V.4.2) La protéine Hfq interagit <i>in vitro</i> avec l'ARNm <i>yeaA</i>	153
V.4.3) La protéine Hfq ne semble pas faciliter l'interaction entre la région 5' de l'ARNm <i>yeaA</i> et l'ARNnc RyhB.	155
V.4.4) Localisation des sites d'interactions mises en place <i>in vitro</i> entre l'ARNnc RyhB, l'ARN <i>yeaA</i> 5' et la protéine Hfq.	156
V.5) Perspectives pour la poursuite de cette étude.	159
Partie 3 : Etude de l'organisation transcriptionnelle du triplet d'ORF <i>epd-pgk-fbaA</i> et de la maturation des transcrits produits.....	
I.) <u>Présentation de l'article.</u>	162
II.) <u>Article II : Characterization of the molecular mechanisms involved in the differential production of Erythrose-4-phosphate dehydrogenase, Phosphoglycerate kinase and class II Fructose-1.6-bisphosphate aldolase in <i>Escherichia coli</i>.</u>	
V. Bardey, C. Vallet, N. Robas, B. Charpentier, B. Thouvenot, E. Hajndorf, P. Régnier, M. Springer and C. Branlant. Soumis à J. Bacteriol.	164
III.) <u>Travaux complémentaires réalisés en vue de déterminer si des évènements post-transcriptionnels interviennent aussi dans la régulation de l'expression du gène <i>gapA</i>.</u>	165
III.1) Test de l'effet du glucose sur les promoteurs <i>gapA</i> P1 et <i>gapB</i> P0 par emploi de fusions transcriptionnelles <i>gapA</i> P1-<i>lacZ</i> et <i>gapB</i> P0-<i>lacZ</i>.	165
III.2) Discussion et perspectives des résultats obtenus.	167
III.3) Le système CsrA/BC n'est pas responsable de l'effet post-transcriptionnel observé sur le promoteur <i>gapA</i> P1.	169

III.3.1) Présentation du système CsrA/BC.....	169
III.3.2) Etude expérimentale de l'effet potentiel de CsrA et CsrB sur l'expression du gène <i>gapA</i>	169
Partie 4 : Analyse de l'effet de l'absence du transporteur spécifique du glucose EIIBC ^{Glc} sur le transcriptome d' <i>E. coli</i>	
I.) <u>Présentation de l'article n°3</u>	171
I.1) Présentation des objectifs	171
I.2) Présentation de la stratégie	172
I.3) Résumé des résultats marquants obtenus	173
II) <u>Article n°3 : An analysis of the consequences of the absence of the EIIBC^{Glc} specific transporter of glucose on porin and flagellin expressions in <i>Escherichia coli</i>.</u> <u>B. Thouvenot, C. Branlant and B. Charpentier</u>	
	174
III.) <u>Discussion</u>	175
III.1) Régulation de la mobilité bactérienne	175
III.2) Régulation de l'expression des porines OmpC et OmpF	176
III.3) Est-ce que les variations du taux de complexe CRP-AMPc et du taux de phosphorylation de la protéine OmpR sont responsables de l'effet du glucose sur les promoteurs <i>gapA</i> P1 et <i>gapB</i> P0 ?	177

Conclusion générale

.....	180
-------	-----

Matériels et Méthodes

Partie 1 : Matériels.....	183
I.) <u>Souches bactériennes utilisées</u>	183
II.) <u>Milieus de culture</u>	184
III.) <u>Plasmides et bactériophages utilisés</u>	185
IV.) <u>Tampons et solutions à usage courant</u>	186
Partie 2 : Méthodes.....	187
I.) <u>Extraction et analyse de l'ADN</u>	187
I.1) Extraction d'ADN plasmidique	187
I.2) Marquage des oligonucléotides	187

I.3) Amplification d'ADN par la technique PCR.....	188
I.4) Clonage de fragments d'ADN.....	188
I.4.1) Ligation d'un fragment d'ADN issu d'une hydrolyse par enzymes de restriction.....	188
I.4.2) Ligation d'un fragment d'ADN synthétisé par la technique PCR.....	189
I.4.3) Vérification de l'insertion.....	189
I.5) Séquençage d'ADN.....	189
I.6) Technique de recombinaison de fragment d'ADN à l'aide du bactériophage λ.....	190
I.6.1) Obtention et production de bactériophages recombinants.....	190
I.6.2) Obtention de souches lysogènes.....	191
I.6.3) Sélection des cellules monolysogènes.....	191
I.7) Mutagenèse dirigée par PCR.....	192
II.) Analyse des ARN.....	193
II.1) Extraction des ARN totaux.....	193
II.2) Technique d'extension d'amorces marquées au ^{32}P par la transcriptase inverse.....	193
II.2.1) Hybridation ARN/oligonucléotide.....	193
II.2.2) Synthèse du brin complémentaire.....	194
II.2.3) Fractionnement des ADNc obtenus.....	194
II.3) Protocoles de transcription <i>in vitro</i>.....	194
II.3.1) Production d'ARN froids ou uniformément marqués par transcription <i>in vitro</i> avec l'ARN pol T7.....	194
II.3.2) Analyse de l'efficacité d'un promoteur par transcription <i>in vitro</i>	195
II.4) Etude des complexes [ARN-Hfq].....	196
II.4.1) Formation des complexes [ARN-Hfq].....	196
II.4.2) Fractionnement des complexes [ARN-Hfq].....	197
II.5) Analyse par la RNase T1.....	197
III) Analyses des protéines cellulaires.....	198
III.1) Cultures et sonication des bactéries.....	198
III.2) Mesure des activités enzymatiques.....	198
III.2.1) Mesure de l'activité β -galactosidase.....	198
III.2.2) Mesures d'activité GAPDH.....	199
III.2.2.1) Préparation du substrat et du cofacteur.....	199
III.2.2.2) Test enzymatique et calcul de l'activité enzymatique.....	200
III.3) Analyse des protéines par gel SDS-PAGE.....	200

Bibliographie

Avant-propos

Au laboratoire, l'équipe de G. Branlant s'est intéressée pendant de nombreuses années à la glycéraldéhyde 3-phosphate déshydrogénase (GAPDH), en tant que modèle d'étude pour le remodelage d'enzyme et en vue de mieux comprendre les bases moléculaires de la spécificité et de la réactivité de cette enzyme. L'équipe de C. Branlant a mené en parallèle une étude de la régulation de l'expression du gène codant cette enzyme chez *Escherichia coli* et d'autres gènes de la glycolyse. Chez *E. coli*, la GAPDH active est codée par le gène *gapA* qui, lorsqu'il est cloné dans un plasmide multicopie, conduit à une très forte production de GAPDH, constituant environ 80 % des protéines solubles (Branlant *et al.*, 1983). L'étude des mécanismes moléculaires expliquant la forte efficacité de la transcription de ce gène présentait donc un intérêt scientifique et biotechnologique. A mon arrivée au laboratoire, B. Charpentier avait montré que cette efficacité transcriptionnelle est liée à la présence de 4 promoteurs régulés différemment qui permettent une expression très forte de ce gène dans un grand nombre de conditions environnementales (Charpentier et Branlant, 1994). Parmi ces 4 promoteurs, le promoteur *gapA* P1 qui est le promoteur le plus efficace en phase exponentielle de croissance n'avait pas, d'après l'examen de la séquence, une architecture qui pouvait expliquer son efficacité. Dans un premier temps, j'ai focalisé mes travaux sur l'étude de l'architecture de ce promoteur. Tout d'abord, par mutagenèse dirigée et par étude de l'efficacité de transcription *in vitro* et *in vivo*, j'ai montré que la structure responsable de l'initiation est de type promoteur étendu avec la particularité que l'activité nécessite la présence d'une boîte -35 de séquence proche de la séquence consensus. Nous avons montré que l'importance de cette boîte est liée au fait que la séquence de la boîte -10 est sous-optimale. De plus, j'ai montré que la courbure de l'ADN générée par deux enchaînements de résidus T en phase avec le pas de l'hélice, situé l'un dans la région interboîte et l'autre dans l'élément -10 du promoteur *gapA* P1, était un déterminant important de l'efficacité de ce promoteur. Nous avons interprété ces données, ainsi que les données d'empreintes chimiques obtenues par B. Charpentier sur la structure du complexe RPo formé à ce promoteur sur la base des données de structure 3D actuellement disponibles dans la littérature. Ces travaux ont fait l'objet d'une publication dans *Biochemical Journal*.

Chez *E. coli*, le gène *yeaA* est situé en amont du gène *gapA* et il est exprimé sur le brin opposé. Il code une protéine ayant une activité Msr (méthionine sulfoxyde réductase). La région inter-ORF *gapA-yeaA* est très complexe. En effet, en plus des 4 promoteurs permettant la transcription du gène *gapA*, deux autres promoteurs qui avaient été mis en évidence par B.

Charpentier permettent d'initier la transcription du gène *yeaA*. Des sites potentiels de fixation de la protéine Fis avaient été mis en évidence, ainsi qu'une forte courbure de l'ADN qui chevauche un site de fixation du complexe CRP-AMPc. B. Charpentier avait montré que la fixation de ce complexe était nécessaire à l'activité du promoteur *gapA* P3. J'ai complété l'étude de cette région inter-ORF en étudiant l'importance des différents éléments structuraux et de fixation des protéines sur la transcription des gènes *gapA* et *yeaA*.

Chez *E. coli*, un second gène *epd*, code une protéine présentant de fortes homologues avec les GAPDH de type bactérien. Néanmoins, le produit de ce gène a une très faible activité GAPDH, mais présente une activité érythrose 4-phosphate déshydrogénase (E4PDH) (Zhao *et al.*, 1995; Boschi-Muller *et al.*, 1997). Le gène *epd* est localisé en amont des gènes *pgk* et *fbaA*, codant respectivement la 3-phosphoglycérate kinase et la fructose 1,6-bisphosphate aldolase, qui interviennent dans la glycolyse et la néoglucogenèse (Alefounder et Perham, 1989). L'idée est que ce gène *epd* ait été généré par mutation du gène *gapA* initial généralement localisé en amont de l'ORF *pgk*, ceci par un nombre limité de mutations (Boschi-Muller *et al.*, 1997; Fillinger *et al.*, 2000). B. Charpentier et V. Bardey avait mis en évidence l'existence d'un promoteur unique, *gapB* P0, localisé en amont du gène *epd* dont l'activité est forte en présence de glucose. Avec C. Vallet, nous avons montré que le promoteur *gapB* P0 permet l'expression des ORF *epd*, *pgk* et *fbaA*. La protéine codée par le gène *epd* est produite en quantité très faible dans la cellule, alors que son gène est transcrit à un taux élevé en présence de glucose dans le milieu de culture. V. Bardey avait montré que les transcrits initiés à ce promoteur sont maturés par la RNase E au niveau de l'ORF *epd*, ce qui conduit à la dégradation spécifique de ces ARNm. Nous avons montré qu'un autre promoteur *pgk* P1 situé en amont de l'ORF *pgk* permet l'expression des ORF *pgk* et *fbaA*. Son efficacité varie peu en fonction des conditions de croissance. Ce promoteur appartenant à la famille des promoteurs étendus doit permettre une expression de base des enzymes PGK et FBA. L'ORF *fbaA* n'a aucun promoteur spécifique. Les enzymes PGK et FBA sont produites de façon différentielle du fait de l'existence d'un processus de maturation dans la région *pgk-fbaA*. L'ensemble de ce travail a fait l'objet d'un article qui est soumis à la revue Journal of Bacteriology.

Les données du laboratoire montraient donc une activation des promoteurs *gapA* P1 et *gapB* P0 en présence de glucose dans le milieu de culture (Gschaedler *et al.*, 1999) et B. Charpentier, V. Bardey et N. Robas avaient montré que l'augmentation du taux de transcrits

initiés à ces deux promoteurs en présence de glucose dépend de l'expression du gène *ptsG* codant la protéine EIIBC^{Glc}, protéine impliquée dans la perméation du glucose (Charpentier *et al.*, 1998).

Pour essayer de comprendre quel était le lien entre l'expression de la protéine EIIBC^{Glc} et l'activité des promoteurs *gapA* P1 et *gapB* P0, nous avons entrepris une étude de l'effet de la mutation du gène *ptsG* sur l'expression des gènes d'*E. coli*. Pour cela, nous avons utilisé une approche globale de transcriptomique réalisée sur les ARNm totaux isolés d'une souche *ptsG*⁻ et comparés aux ARNm isolés d'une souche *ptsG*⁺. Les résultats montrent que l'absence de protéine EIIBC^{Glc} active entraîne un remaniement fort du taux d'expression de certains gènes, notamment ceux codant des constituants de la membrane externe, des constituants des flagelles et des enzymes du métabolisme carboné. Nous avons ainsi montré que les souches *ptsG*⁻ étaient plus mobiles et que la membrane externe était plus perméable du fait de la présence majoritaire de la porine OmpF qui permet la formation de pores de diamètre plus important. Un article présentant les résultats de cette étude a été préparé.

Nous commencerons donc par faire une revue bibliographique sur le mécanisme d'initiation de la transcription chez *E. coli* en insistant sur les bases moléculaires qui permettent l'ouverture de la double hélice d'ADN par l'ARN polymérase. Ensuite, nous décrirons les régulations de l'initiation de la transcription exercées par des facteurs agissant en *cis* (séquence des promoteurs et influence de la structure de l'ADN) ou en *trans* (protéines qui se fixent sur l'ADN) qui permettront d'expliquer les régulations observées au niveau de la région inter-ORF *gapA-yeaA*. Enfin, nous terminerons l'introduction bibliographique en présentant les systèmes qui permettent l'adaptation d'*E. coli* aux changements environnementaux. Ces problématiques seront importantes pour la compréhension de nos données expérimentales obtenues lors de l'étude de l'effet de l'absence du transporteur EIIBC^{Glc} par l'approche puces à ADN. Les résultats seront présentés sous forme d'article publié, soumis ou à être soumis, avec pour chacun une présentation des objectifs et une discussion incorporant certaines données non insérées dans les articles. Un bilan et une discussion générale fera suite à cette partie Résultats.

Introduction

Partie 1 : La transcription chez *Escherichia coli*.

La transcription de l'information génétique dans toutes les cellules vivantes est réalisée par des ARN polymérase ADN dépendantes (ARN pol) qui comportent un nombre de sous-unités différent chez les bactéries, les Archaea et chez les eucaryotes. Chez les eucaryotes, trois ARN pol différentes sont nécessaires à la transcription de tous les gènes, alors que chez les bactéries et les Archaea, une seule enzyme assure la transcription de tous les gènes. Les sous-unités catalytiques des ARN pol des bactéries, des chloroplastes, des mitochondries et des ARN pol nucléaires font partie d'une même famille structurale de protéines (pour revue : Ebright, 2000).

Comprendre le mécanisme d'action des ARN pol bactériennes apporte donc des informations sur le mécanisme d'action des polymérase eucaryotes plus compliquées, en particulier au niveau du nombre de sous-unités.

La transcription se décompose en trois étapes :

- l'initiation.
- l'élongation.
- la terminaison.

Chez les bactéries, lors de l'étape d'initiation de la transcription, l'ARN pol est associée à un facteur σ pour former une holoenzyme $E\sigma$, alors que l'étape d'élongation est assurée par l'apoenzyme qui possède l'activité catalytique de synthèse d'ARN.

Le premier contrôle de l'expression des gènes s'effectue très souvent à l'étape d'initiation de la transcription. En effet, de nombreux mécanismes de régulations reposent sur l'activation ou la répression de cette étape. Des éléments ADN et les protéines régulatrices qui s'y fixent permettent la sélection des gènes à transcrire et définissent le taux de synthèse des transcrits. Cette régulation à une étape précoce de l'expression des gènes permet une économie d'énergie à la cellule.

Sur le génome bactérien, la transcription est initiée au niveau de régions précises de l'ADN situées en amont des gènes à transcrire. Ces séquences appelées promoteurs sont les sites de reconnaissance pour l'ARN pol.

L'initiation de la transcription se décompose elle-même en plusieurs sous-étapes. La première sous-étape est la fixation de l'ARN pol sur la séquence promotrice. Cette étape est essentielle puisqu'elle permet la sélection du gène qui sera transcrit. Une fois fixée, l'ARN pol va permettre l'ouverture de la double hélice et le complexe ARN pol-promoteur, ainsi formé, est appelé RPo (R = ARN pol, P = promoteur, o = open). L'étape suivante est l'enchaînement des premiers nucléotides de l'ARN. Après relargage du facteur σ , l'ARN pol va ensuite réaliser l'étape d'élongation pour permettre la synthèse des transcrits de grande taille. La dernière étape est l'arrêt de la transcription appelée terminaison.

Chacune des étapes de l'initiation peut représenter une cible de régulation. Ces régulations font intervenir plusieurs paramètres, notamment la séquence et la structure de l'ADN au niveau du promoteur et celle des régions régulatrices avoisinantes, ainsi que des effecteurs protéiques.

I.) Le mécanisme de l'initiation de la transcription chez les bactéries.

I.1) Les acteurs.

Nous commencerons par décrire les caractéristiques des séquences ADN sur lesquelles se fixent les ARN pol, à savoir, les promoteurs.

Chez *E. coli*, 7 facteurs σ peuvent se fixer sur l'apoenzyme. Les 7 holoenzymes ainsi formées ont des spécificités différentes de reconnaissance de l'ADN. Ces séquences reconnues par chaque holoenzyme sont décrites dans le tableau 1. La majorité des gènes sont transcrits par l'holoenzyme associée au facteur σ^{70} . Nous décrivons donc plus en détail les séquences reconnues par l'holoenzyme $E\sigma^{70}$.

I.1.1) Les promoteurs.

A partir de la comparaison d'un grand nombre de séquences permettant l'initiation de la transcription par l'holoenzyme $E\sigma^{70}$, une séquence consensus a été établie (Fig. 1) (Harley et Reynolds, 1987; Lissner et Margalit, 1993; Ozoline *et al.*, 1997; Mitchell *et al.*, 2003). La même approche a permis de définir les séquences consensus des promoteurs reconnus par les holoenzymes renfermant d'autres facteurs σ (Tableau 1).

Facteur	Gène	Masse en Da	Séquence consensus des promoteurs		Type de gènes transcrits par l'holoenzyme	Famille de facteurs (groupe)
			boîte -35	boîte -10		
σ^{70}	<i>rpoD</i>	70236	TTGACA N16-18	TATAAT	Majorité des gènes d' <i>E. coli</i>	σ^{70} (primaire)
$\sigma^{38(S)}$	<i>rpoS</i>	38000	<u>TTGACA</u> N15-17	CTATACT	Gènes exprimés lors de la phase stationnaire et lors de nombreux stress	σ^{70} (primaire non essentiel)
σ^{32}	<i>rpoH</i>	32000	CTTGAA N13-15	GNCCCCATNT	Gènes de réponse à un stress entraînant la dénaturation des protéines cytoplasmiques	σ^{70} (alternatif)
$\sigma^{28(F)}$, FliA	<i>rpoF</i>	28000	TAAA N15	GCCGATAA	Gènes codant la synthèse des flagelles	σ^{70} (alternatif)
$\sigma^{24(E)}$	<i>rpoE</i>	24000	GAACCTT N16	TCTGA	Gènes de réponse à un stress entraînant la dénaturation des protéines périplasmiques	σ^{70} (ECF)
σ^{19} (FecI)	<i>fecI</i>	19000	GGAAAAT N17	TCCTTT	Opéron de transport du citrate ferrique	σ^{70} (ECF)
$\sigma^{54(N)}$	<i>rpoN</i>	54000	CTGGCAC N5	TTGC	Gènes du métabolisme de l'azote	σ^{54}
			-26	-12		

Tableau 1 : Les différents facteurs sigma mis en évidence chez *E. coli* et les séquences consensus établies pour les promoteurs transcriptionnels reconnus par l'ARN pol contenant ces facteurs (d'après Helmann et Chamberlin, 1988 ; Wosten, 1998).

Pour le facteur σ^{54} , les éléments de reconnaissance se situent entre les positions -12 et -26 par rapport au site d'initiation de la transcription. La séquence de la boîte -35 est soulignée pour le facteur σ^{38} car cette séquence n'est pas obligatoire pour la reconnaissance par l'ARN pol $E\sigma^S$.

Les familles auxquelles appartiennent les facteurs σ sont indiquées dans la dernière colonne. Il existe deux types de facteurs σ chez *E. coli* : les facteurs homologues au facteur σ^{70} et le facteur σ^{54} . Dans la famille σ^{70} , quatre sous-familles sont représentées (primaires, primaires non essentiels, alternatifs et ECF).

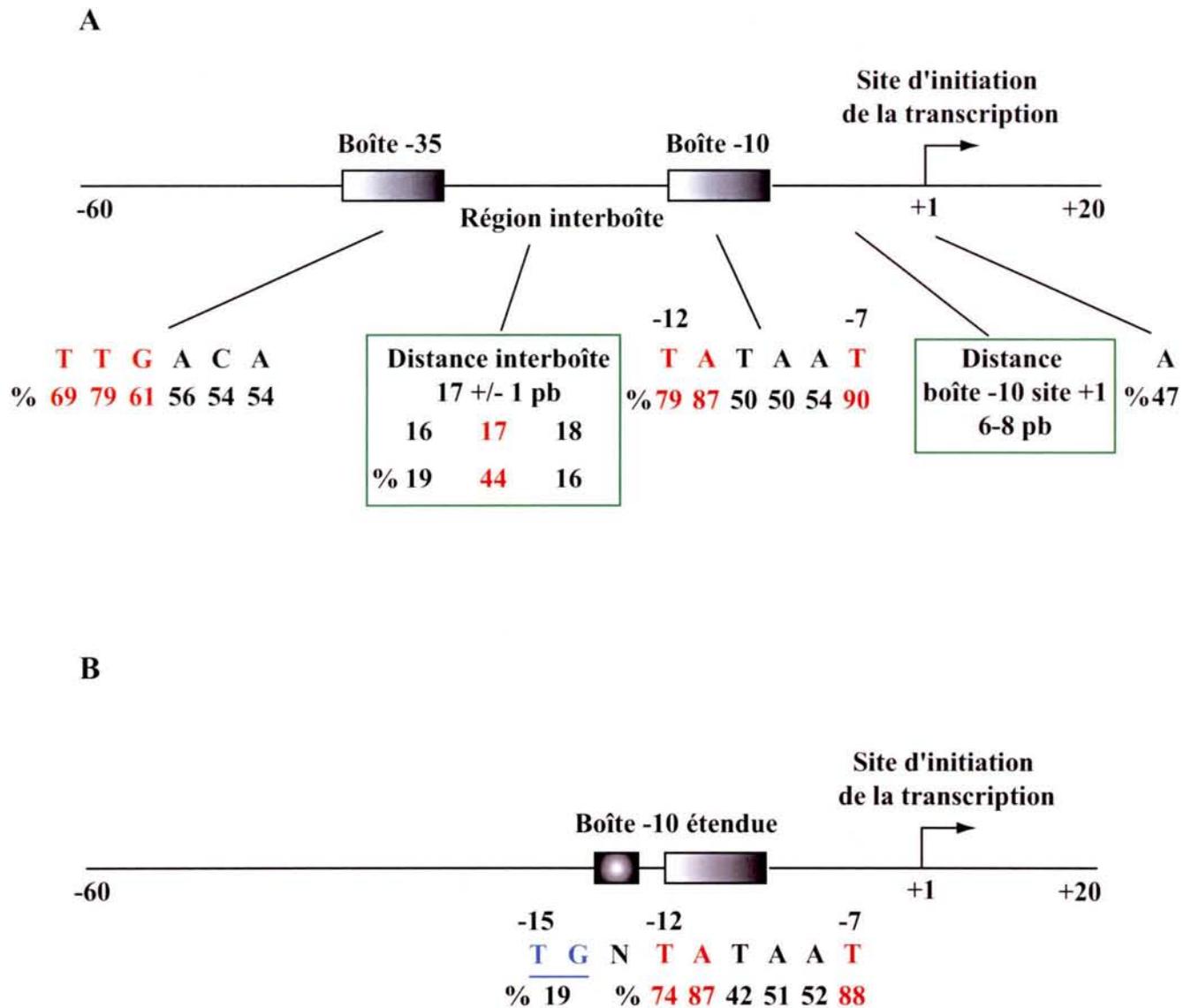


Figure 1 : Séquences consensus établie pour les promoteurs reconnus par l'ARN pol $E\sigma^{70}$ chez *E. coli*.

A) Séquences consensus des promoteurs "classiques" possédant une boîte -10 et une boîte -35. Le site d'initiation de la transcription (+1) est indiqué par une flèche. Les séquences consensus et les distances optimales entre ces séquences sont indiquées en dessous du schéma. Pour chaque position, les préférences nucléotidiques sont indiquées par le pourcentage de présence du nucléotide indiqué, déterminé par comparaison des séquences de différents promoteurs d'*E. coli* (Lisser et Margalit, 1993 et Mitchell *et al.*, 2003). Les positions et distances les plus conservées sont indiquées en rouge.

B) Séquence consensus des promoteurs étendus. Le site d'initiation de la transcription (+1) est indiqué par une flèche. Un promoteur étendu est caractérisé par la présence d'un dinucléotide TG situé une paire de base en amont de la boîte -10 (positions -14 et -15 par rapport au +1). Ce motif TG est présent dans 19% des promoteurs d'*E. coli* (Mitchell *et al.*, 2003). Les positions en rouge indiquent les positions les plus conservées.

Dans le cas de l'holoenzyme $E\sigma^{70}$, la séquence consensus est caractérisée par la présence de deux séquences hexamériques appelées boîtes -10 et -35.

- La séquence consensus de la boîte -10 est 5'-TATAAT-3'. Le premier nucléotide est, dans un grand nombre de cas, situé à la position -12 par rapport au site d'initiation de la transcription (+1). Ainsi, le dernier nucléotide est séparé du site +1 par 6 pb.
- La séquence de la boîte -35 est 5'-TTGACA-3'. Le résidu 3' de cette boîte est séparé par 17 +/- 1 pb de la boîte -10.

La séquence située entre les boîtes -10 et -35 est appelée région interboîte.

Les deux bases les plus conservées dans la boîte -10 sont le T en position 1, avec une fréquence de 79%, le A en position 2 (87%) et le T en position 6 (90%). Les trois autres positions sont moins conservées avec des pourcentages de présence de l'ordre de 50% (Mitchell *et al.*, 2003) (Fig. 1).

Le niveau de conservation de l'identité des bases de la boîte -35 est moins importante que celui des résidus de la boîte -10. Les bases les plus conservées dans ce motif sont TTA avec des pourcentages de présence de respectivement, 69, 79 et 61% (Fig. 1).

Dans 44% des promoteurs étudiés, la distance interboîte est de 17 pb et pour 79% d'entre-eux, elle est comprise entre 16 et 18 pb.

Certains promoteurs sont fonctionnels alors que la séquence située 17 paires de bases en amont de la boîte -10 n'a que très peu d'homologie avec la séquence consensus définie pour les boîtes -35. Pour ces promoteurs, la présence d'un site de fixation d'un activateur de la transcription pourrait remplacer fonctionnellement la boîte -35 (des exemples seront décrits dans la partie régulation de la transcription).

A la fin des années 1980, un autre type de promoteur reconnu par l'holoenzyme $E\sigma^{70}$ appelé promoteur étendu (pour revue : Bown *et al.*, 1997), a été mis en évidence (Fig. 1 B). Ces promoteurs sont caractérisés par la présence d'un dinucléotide TG situé une paire de base en amont de la boîte -10 (c'est à dire aux positions -14 et -15 par rapport au site +1 d'initiation de la transcription) (Keilty et Rosenberg, 1987; Burns et Minchin, 1994; Chan et Busby, 1989; Bown *et al.*, 1997). La séquence consensus des boîtes -10 des promoteurs étendus est donc : 5'TGNTATAAT3' (Keilty et Rosenberg, 1987). Il a été proposé que le dinucléotide TG soit un autre site de reconnaissance pour la polymérase et que cette interaction compense l'absence de la boîte -35 (Bown *et al.*, 1997). En effet, pour les premiers promoteurs étendus étudiés, la séquence située au niveau de la boîte -35 ne présentait pas d'homologie avec la séquence consensus des boîtes -35 (Keilty et Rosenberg, 1987; Kumar *et al.*, 1993). Par

ailleurs, une analyse exhaustive récente des promoteurs d'*E. coli* montre qu'un certain nombre de promoteurs présentant un dinucléotide TG pourraient avoir au moins 3 pb sur 6 identiques au consensus pour la boîte -35 (Mitchell *et al.*, 2003). Par contre, l'influence de la séquence de cette boîte -35 sur des promoteurs étendus naturels n'avait pas été testée expérimentalement.

Les promoteurs étendus représentent environ 20% de l'ensemble des promoteurs d'*E. coli* reconnu par σ^{70} (Burr *et al.*, 2000; Mitchell *et al.*, 2003). Chez d'autres bactéries, les promoteurs étendus sont beaucoup plus fréquents, notamment chez les bactéries à Gram positif, où ce type de promoteur est présent à plus de 45% (Helmann, 1995).

Par ailleurs, chez les bactéries Gram⁺, notamment chez *B. subtilis*, une conservation des résidus en amont du dinucléotide TG est même observée. Ainsi, le motif TRTG (ou R représente une purine) est très important pour une efficacité optimale du promoteur *amyP* (Voskuil *et al.*, 1995; Voskuil et Chambliss, 2002).

Burr *et al.* (2000) ont déterminé l'efficacité de l'initiation de la transcription de promoteurs étendus d'*E. coli*, en introduisant toutes les combinaisons possibles de résidus aux positions -17, -16, -15 et -14 (qui comprend le motif TG). Cette étude a montré que, comme pour *B. subtilis*, l'identité des bases en amont du motif TG est importante et l'activité la plus élevée est obtenue avec la séquence TGTG. L'effet de la présence de cette séquence est surtout visible pour des promoteurs dont la séquence des boîtes -10 et -35 est éloignée du consensus.

Contrairement aux promoteurs classiques, certains promoteurs étendus comme le promoteur *galPI* peuvent initier la transcription à basse température (Burns *et al.*, 1996) mais il ne semble pas que cette caractéristique soit commune à tous les promoteurs étendus. En effet, le promoteur *PcysG* qui est un promoteur étendu ne permet pas l'initiation de la transcription à des températures inférieures à 37°C (Belyaeva *et al.*, 1993).

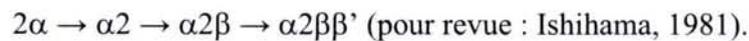
Les séquences reconnues par les autres holoenzymes sont différentes mais toujours constituées de deux boîtes dont les positionnements par rapport au site d'initiation de la transcription varient légèrement suivant les enzymes. Dans le cas du facteur σ^{54} , les boîtes qui permettent la fixation de l'holoenzyme sont appelées boîtes -26 et -12 (Tableau 1).

Après avoir décrit les séquences permettant la fixation de l'holoenzyme, nous allons présenter les analyses structurales et fonctionnelles sur les enzymes bactériennes.

I.1.2) L'ARN polymérase bactérienne (EC 2.7.7.6).

I.1.2.1) L'apoenzyme.

La forme apoenzyme est composée de cinq sous-unités différentes : deux sous-unités α , une sous-unité β , une sous-unité β' et une sous-unité ω (Tableau 2). Ces cinq sous-unités s'associent suivant un ordre précis pour former l'apoenzyme :



Sous-unité	Gène	Masse en Da	Taille en aa	Référence	Structure et fonctions
α	<i>rpoA</i>	36512	329	(Ovchinnikov <i>et al.</i> , 1977)	Constituée de deux domaines (Blatter <i>et al.</i> , 1994). Initiation de l'assemblage de l'apoenzyme (Ishihama, 1981) par dimérisation du domaine N-terminal (Igarashi <i>et al.</i> , 1991a).
β	<i>rpoB</i>	150612	1342	(Ovchinnikov <i>et al.</i> , 1981)	Possède l'activité catalytique de synthèse de l'ARN, ainsi que l'activité de correction des erreurs (Record <i>et al.</i> , 1996).
β'	<i>rpoC</i>	155162	1407	(Ovchinnikov Yu <i>et al.</i> , 1982)	Formation du site catalytique en association avec la sous-unité β (Record <i>et al.</i> , 1996).
ω	<i>rpoZ</i>	10105	90	(Burgess <i>et al.</i> , 1969)	Activité chaperone pour le repliement de la sous-unité β' (Ghosh <i>et al.</i> , 2001)

Tableau 2 : Caractéristiques des sous-unités constituant la forme apoenzyme de l'ARN pol d'*E. coli*.

Cette apoenzyme possède une activité catalytique permettant la formation des liaisons phosphodiester 3'-5' liant les ribonucléotides. La structure de l'apoenzyme d'*Escherichia coli* a été établie à une résolution de 15 Å par cryo-microscopie électronique. La structure ainsi obtenue est dite en « pince de crabe » (Darst *et al.*, 2002) (Fig. 2 B(1)). D'autres structures ont été établies mais avec des résolutions beaucoup plus faibles (Darst *et al.*, 1989; Polyakov *et al.*, 1995; Darst *et al.*, 1998).

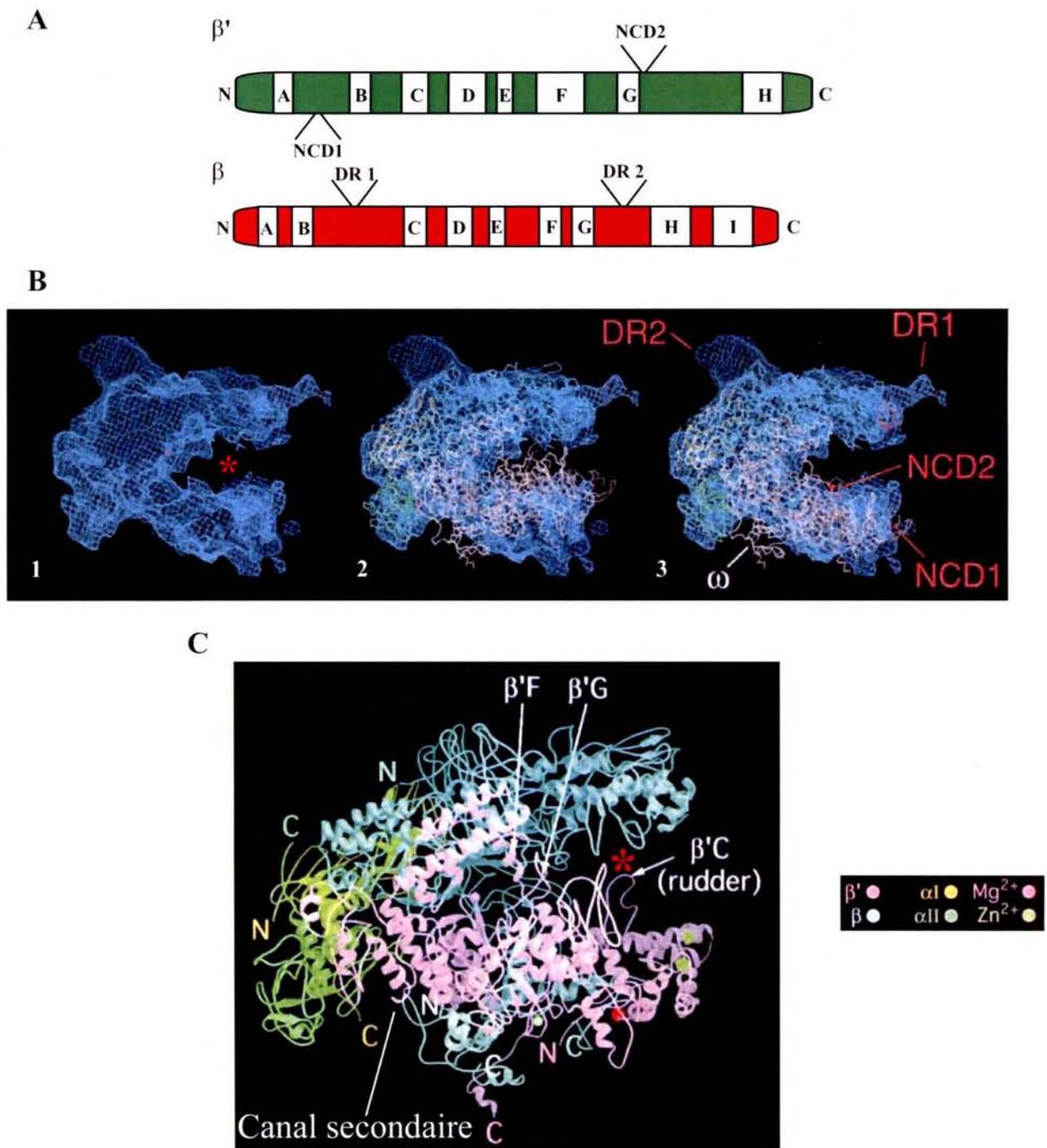


Figure 2 : Analyse de la structure des ARN pol d'*E. coli* et *T. aquaticus* sous leur forme apoenzyme (d'après Darst *et al.*, 2002 et Zhang *et al.*, 1999).

A) Représentation schématique des sous-unités β et β' de l'ARN pol. Les régions notées de A à H pour β et A à I pour β' sont les régions conservées chez les procaryotes. La présence d'insertions ou de délétions entre les apoenzymes d'*E. coli* et de *T. aquaticus* est indiquée : en haut pour des insertions dans la séquence d'*E. coli* (NCD2, DR1 et DR2) et en bas pour des insertions dans la séquence de *T. aquaticus* (NCD1).

B) Comparaison des structures tridimensionnelles des apoenzymes d'*E. coli* et *T. aquaticus* (d'après Darst *et al.*, 2002) (1) densités électroniques de l'apoenzyme d'*E. coli*, (2) superposition des structures des apoenzymes d'*E. coli* et *T. aquaticus* et (3) l'écart entre les deux "pinces" étant différent pour les deux apoenzymes, les auteurs ont effectué un déplacement vectoriel des "pinces" de l'apoenzyme de *T. aquaticus* pour que les deux structures se superposent. Les différentes insertions, ainsi que la présence de la sous-unité ω , sont indiquées. NCD1/2 : nonconserved domain 1/2 et DR1/2 : dispensable region 1/2. * indique la position du canal de fixation de l'ADN.

C) Structure de l'apoenzyme de *T. aquaticus* (Zhang *et al.*, 1999). Les couleurs utilisées pour la représentation des sous-unités et des ions métalliques sont indiquées à droite de la structure. Les extrémités N et C terminales sont indiquées pour chaque sous-unité. Les éléments structuraux cités dans le texte sont indiqués. * indique la position du canal de fixation de l'ADN.

Ces données sur l'enzyme d'*E. coli* ont été comparées aux données obtenues par diffraction aux rayons X à 3,3 Å de résolution pour l'apoenzyme bactérienne de l'hyperthermophile *Thermus aquaticus* (Fig. 2 C) (Zhang *et al.*, 1999; Darst *et al.*, 2002). Les deux enzymes présentent une forme similaire en « pince de crabe », les sous-unités β et β' constituant les pinces (Darst, 2001; Darst *et al.*, 2002). Les dimensions des deux complexes sont d'environ 150 Å de long, 110 Å de large et 115 Å de haut. La différence majeure entre les deux complexes est l'écartement des deux « pinces » qui est plus grand pour l'apoenzyme d'*E. coli* (Fig. 2 B(3) (Darst *et al.*, 2002)).

Lorsque les séquences primaires des sous-unités β et β' des enzymes d'*E. coli* et de *T. aquaticus* sont comparées, l'existence de plusieurs insertions sont mises en évidence (Fig. 2 A). Elles se situent dans des régions non essentielles pour la fonction de l'ARN pol et constituent des domaines autonomes dont la délétion n'affecte pas la structure générale de l'apoenzyme (Darst *et al.*, 2002). Ces insertions présentes soit dans l'ARN pol d'*E. coli* (DR1, DR2 [dispensable region 1 et 2] et NCD2 [nonconserved domain 2]), soit dans l'ARN pol de *T. aquaticus* (NCD1 nonconserved domain 1) pourraient intervenir dans des régulations de l'activité enzymatique (Darst *et al.*, 2002) ou n'avoir aucune fonction mais être non gênantes pour l'activité.

Les structures des enzymes d'*E. coli* et de *T. aquaticus* renferment un canal qui doit permettre d'accommoder la double hélice d'ADN servant de matrice pour la synthèse d'ARN. Pour la structure de l'apoenzyme de *T. aquaticus*, la largeur de ce canal est de 27 Å (Zhang *et al.*, 1999) tandis que pour l'apoenzyme d'*E. coli*, la largeur est plus importante (50 Å) (Darst *et al.*, 2002). Le site actif de l'enzyme est situé au fond de ce canal. Ce site actif comporte un ion Mg^{2+} essentiel à la catalyse. Un deuxième ion Mg^{2+} semble aussi essentiel à la catalyse mais sa position n'a pas été définie dans la structure de l'apoenzyme. Une structure formée par une région de la sous-unité β' localisée au milieu du canal (β' F et G, Fig. 2) subdivise ce canal en un canal principal et un canal secondaire qui pourrait permettre aux nucléotides de pénétrer à l'intérieur de l'enzyme afin d'atteindre le site actif (Darst, 2001). Un autre élément structural important pour la fonction de l'ARN polymérase appelé « levier » β' C (rudder) (Fig. 2 C) semble impliqué dans le maintien de l'ouverture de la double hélice d'ADN (Darst, 2001). Plusieurs autres rôles sont possibles pour cet élément structural du fait de sa proximité avec l'hybride ADN-ARN formé lors de la réaction (Korzheva *et al.*, 2000; Kuznedelov *et al.*, 2002a; Cramer, 2002). Ainsi, il a été proposé que cet élément favorise la séparation de l'ARN néosynthétisé de l'ADN matrice participant ainsi au maintien de la taille de l'hybride ADN-

ARN ou d'interagir directement avec l'ARN, ce qui stabiliserait le complexe d'élongation (Korzheva *et al.*, 2000; Kuznedelov *et al.*, 2002a). *In vitro*, l'ARN pol de *T. aquaticus* dans laquelle cette région a été déléetée ne semble pas affectée au niveau des capacités de séparation et de maintien de l'hybride. Par contre, la stabilité du complexe d'élongation semble affectée, ce qui suggère que cet élément structural interagit directement avec l'ARN (Kuznedelov *et al.*, 2002a). La sous-unité β présente une extension mobile, appelée « clapet » β (flap) dont la structure et la dynamique semble indépendante du reste de la sous-unité (Fig. 2 C). Elle serait impliquée dans la terminaison de la transcription (Darst, 2001), dans les pauses de l'ARN pol en cours d'élongation (Toulokhonov et Landick, 2003) et aussi dans la reconnaissance des séquences du promoteur (Kuznedelov *et al.*, 2002b).

Le rôle potentiel des différentes sous-unités de l'ARN pol a pu être déterminé en employant des méthodes très variées comme la production et l'étude des propriétés de mutants affectés dans la fonction de l'enzyme ou l'utilisation de sondes photoréactives fixées aux extrémités des transcrits naissants. Ces données ont été exploitées d'un point de vue structural, grâce à la connaissance de la structure de l'apoenzyme de *T. aquaticus* (Zhang *et al.*, 1999). La figure 3 montre les régions conservées des sous-unités qui composent l'apoenzyme. Il existe 2 régions conservées pour α , 9 pour β (A-I) et 8 pour β' (Allison *et al.*, 1985; Sweetser *et al.*, 1987).

Régions et sous-unités impliquées dans l'assemblage de l'apoenzyme et de l'holoenzyme.

Dans l'ARN pol, la sous-unité α est présente en deux exemplaires. Elle est surtout impliquée dans l'assemblage de l'apoenzyme. Elle est constituée de deux domaines (Blatter *et al.*, 1994) séparés par une boucle flexible de 13 résidus (Fig. 3) (Jeon *et al.*, 1997). Le domaine N-terminal (acides aminés de 1 à 236) est surtout impliqué dans la dimérisation des deux sous-unités et l'assemblage de l'ARN pol (Igarashi *et al.*, 1991a; Heyduk *et al.*, 1996), tandis que le domaine C-terminal (acides aminés de 249 à 329) est surtout impliqué dans les phénomènes d'activation de la transcription, par interaction avec des complexes protéiques (cf chapitre) (Hayward *et al.*, 1991; Igarashi et Ishihama, 1991c; Tang *et al.*, 1994) ou des séquences d'ADN appelées élément « UP » (cf chapitre) (Ross *et al.*, 1993; Blatter *et al.*, 1994). Le domaine N-terminal se décompose en deux parties. La région portant les acides aminés 30 à 75 est impliquée dans la fixation à la sous-unité β , alors que celle comprenant les acides aminés 175 à 210 est impliquée dans la liaison à la sous-unité β' (Heyduk *et al.*, 1996).

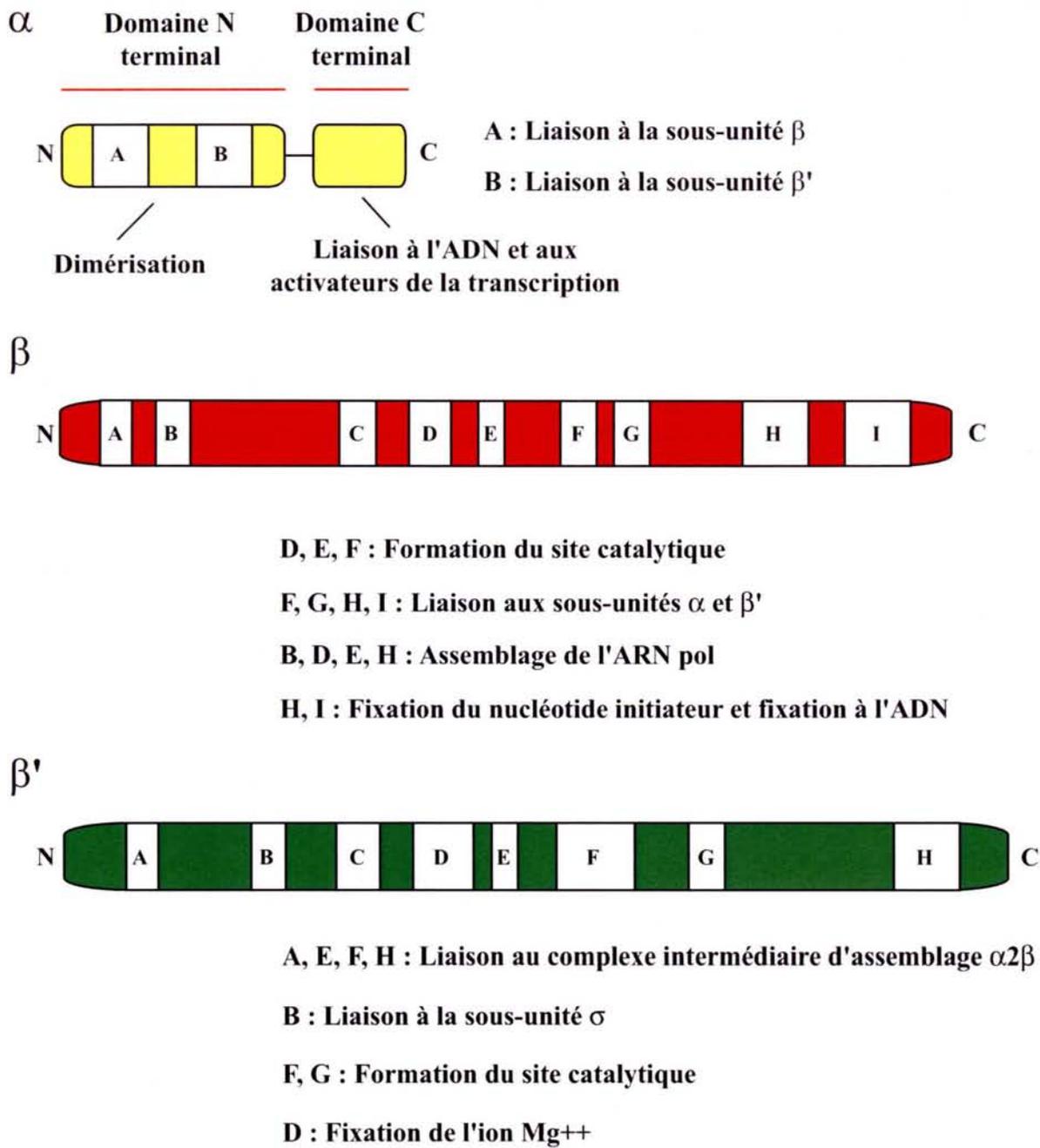


Figure 3 : Représentation schématique des différentes sous-unités de l'ARN pol d'*E. coli*. Les régions notées de A à B pour α , A à I pour β et A à H pour β' sont les régions les plus conservées parmi les ARN pol d'archaea, de bactéries et d'eucaryotes. Les différentes fonctions de ces zones conservées sont indiquées.

Ces régions d'interaction sont les régions les plus conservées dans la séquence de la sous-unité α (Fig. 3) (Heyduk *et al.*, 1996).

La région 260 à 309 (région conservée B) de la sous-unité β' est impliquée dans la liaison au facteur σ (Arthur et Burgess, 1998). D'autres régions de β' entrent en interaction avec des résidus de la sous-unité σ mais les interactions sont plus faibles (Katayama *et al.*, 2000). Les autres régions conservées de β' sont impliquées dans la liaison à l'intermédiaire d'assemblage $\alpha 2\beta$ (régions A, E, F et H) (Fig. 3) (Arthur *et al.*, 2000; Katayama *et al.*, 2000). Au sein de la sous-unité β , les domaines conservés D et E sont, respectivement, à proximité des domaines conservés B et H (Tavormina *et al.*, 1996). Ils sont importants pour l'assemblage de l'ARN pol. Les acides aminés qui se situent dans les domaines conservés H et I sont aussi très importants pour l'assemblage de l'apoenzyme. En effet, les acides aminés de ces 2 régions sont impliqués dans l'interaction avec les sous-unités α et β' . De plus, les régions conservées F et G participent à ces interactions (Fig. 3) (Wang *et al.*, 1997a).

Le gène *rpoZ* qui code le facteur ω n'est pas essentiel à la viabilité des cellules d'*E. coli*. Le gène codant cette sous-unité est pourtant présent chez un grand nombre de génomes bactériens séquencés, ce qui indique une fonction conservée et importante. Le facteur ω ne serait pas nécessaire à l'activité de l'ARN pol mais aurait plutôt un rôle structural (Mukherjee et Chatterji, 1997). Il se fixe à la sous-unité β' (Mukherjee et Chatterji, 1997; Ghosh *et al.*, 2003) et facilite la liaison des extrémités N et C terminales, ce qui permet de maintenir une conformation de β' compatible avec son association à l'intermédiaire $\alpha 2\beta$ (Ghosh *et al.*, 2001) La sous-unité ω joue donc le rôle de protéine chaperone pour la structuration de la sous-unité β' et favorise donc l'assemblage de l'ARN polymérase.

Régions impliquées dans la formation du site actif.

Le site catalytique de l'ARN pol est porté par les sous-unités β et β' . En remplaçant l'ion catalytique Mg^{++} du site actif de l'ARN pol par un Fe^{++} , Zaychikov *et al.*, (1996) ont obtenu des coupures protéolytiques par action de radicaux hydroxyles, générés par l'ion Fe^{++} . L'analyse de la position des coupures dans les sous-unités β et β' a permis d'identifier les acides aminés du site actif, notamment les trois résidus Asp de la séquence NADFDGD conservée dans le domaine D de β' qui permettent la coordination de l'ion Mg^{++} catalytique (Zaychikov *et al.*, 1996). Un affinement de l'analyse par Mustaev *et al.* (1997) a permis de

mettre en évidence l'implication d'autres acides aminés de la sous-unité β , ainsi que des acides aminés de la sous-unité β' . En ce qui concerne la sous-unité β , ces coupures se situent au niveau des régions conservées D, E, F et H et pour la sous-unité β' , au niveau des régions conservées D, F et G (Fig. 3) (Mustaev *et al.*, 1997). Le site actif de l'enzyme a donc une organisation modulaire. La présence de ces acides aminés dans le site catalytique est en accord avec les données cristallographiques (Zhang *et al.*, 1999; Vassylyev *et al.*, 2002).

Les sous-unités β et β' forment un tunnel qui fixe l'ADN double brin (des positions +5 à +20 par rapport au site +1 d'initiation de la transcription). La présence de contacts entre ce tunnel et l'ADN est en accord avec des données obtenues par l'étude de l'empreinte de l'enzyme sur l'ADN (Schickor *et al.*, 1990; Craig *et al.*, 1995). L'interaction de l'ADN avec ce tunnel semble avoir des rôles multiples. En effet, les interactions ADN (+5 à +20)-ARN pol (β et β') favoriseraient l'ouverture de la double hélice (Roy *et al.*, 2004) mais aussi stabiliseraient le complexe ouvert RPo (Ederth *et al.*, 2002) et pourraient améliorer l'efficacité de l'étape d'élongation (Ederth *et al.*, 2002).

Des études réalisées avec un analogue de nucléotide montrent que des acides aminés de la sous-unité β sont en contact direct avec le nucléotide initiateur (Mustaev *et al.*, 1991). Ces acides aminés se situent dans les domaines conservés H et I de β .

I.1.2.2) L'holoenzyme.

Pour initier la transcription, l'apoenzyme doit interagir avec un des facteurs σ qui confère à l'holoenzyme formée, la spécificité de reconnaissance de régions promoteurs. Ce sont en effet les facteurs σ qui régissent cette spécificité. L'association entre les différents facteurs sigma et l'apoenzyme est très spécifique et fait l'objet de régulations.

I.1.2.2.1) Les facteurs σ .

Les facteurs σ peuvent être classés en deux familles distinctes du point de vue structural et fonctionnel : la famille σ^{70} et la famille σ^{54} (Gruber et Gross, 2003).

La présence de plusieurs facteurs de la famille σ^{70} et d'un représentant de la famille σ^{54} est retrouvé chez beaucoup d'espèces bactériennes mais le facteur σ^{54} n'est pas présent chez les bactéries Gram+ à fort pourcentage en G-C, les mycoplasmes, les mycobactéries et chez les cyanobactéries (Studholme et Buck, 2000). La plupart des bactéries possèdent plusieurs

représentants de la famille σ^{70} . Par exemple, *Streptomyces coelicolor* (Gram⁺) possède 63 facteurs σ de la famille σ^{70} .

La famille σ^{70} a été subdivisée en plusieurs groupes sur la base de la séquence protéique des facteurs σ (Fig. 4):

1) Les facteurs primaires : ce groupe contient le facteur σ^{70} d'*E. coli* (*Eco rpoD*), ainsi que ses orthologues. Les facteurs σ de ce groupe sont essentiels à la croissance. Ils permettent la transcription de la majorité des gènes en phase active de croissance.

2) Les facteurs primaires non essentiels : la séquence des facteurs σ de ce groupe est proche de celle des facteurs primaires mais ces facteurs ne sont pas essentiels à la croissance. Un des exemples est le facteur σ^S d'*E. coli* (*Eco rpoS*), qui est synthétisé en phase stationnaire.

3) Les facteurs alternatifs : la séquence protéique des facteurs est très divergente par rapport à celle des facteurs primaires. Ces facteurs σ contrôlent l'expression de régulateurs qui sont activés lors de conditions très spécifiques de croissance comme un stress thermique, une transition de phase de croissance ou un changement morphologique (Gruber et Gross, 2003). Un exemple de cette famille est le facteur σ^{32} d'*E. coli* (*Eco rpoH*), qui est produit lorsque des protéines sont dénaturées dans le cytoplasme par un stress ou le facteur $\sigma^{28(F)}$ d'*E. coli* (*Eco FliA*), qui permet la synthèse des flagelles (cf. Partie 2).

4) Les facteurs ECF (Extra Cytoplasmic Function) : la séquence protéique de ces facteurs est aussi très divergente par rapport à celle des facteurs primaires. Ces facteurs sont impliqués dans la transcription de gènes codant des protéines ayant des fonctions très variées (Helmann, 2002). Cette famille est très diversifiée et pour certains microorganismes, ces facteurs peuvent être présents en grand nombre. Par exemple, pour *S. coelicolor*, parmi les 63 facteurs σ , 49 facteurs font partie de cette famille (Paget et Helmann, 2003). Deux représentants de cette famille sont présent chez *E. coli*. Il s'agit du facteur σ^{FecI} (*Eco FecI*), impliqué dans l'assimilation du citrate de fer et du facteur σ^E (*Eco rpoE*) qui est produit lorsque des protéines sont dénaturées dans le périplasme en cas de stress (pour revue : Helmann, 2002).

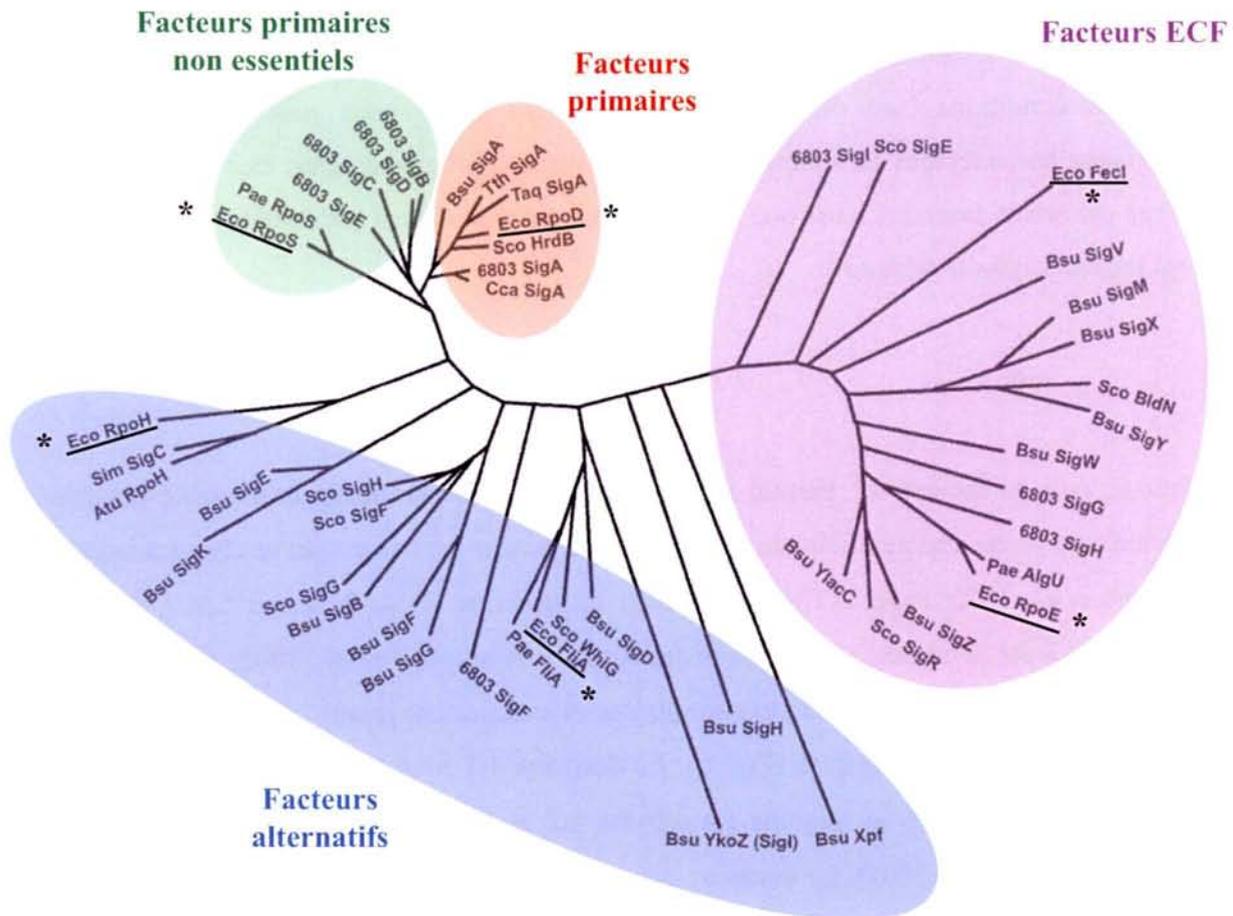


Figure 4 : Arbre phylogénique des facteurs σ de la famille des facteurs σ^{70} (d'après Gruber et Gross, 2003).

Les groupes sont représentés par des couleurs distinctes. Les * indiquent les facteurs σ présents chez *E. coli*.

Eco, *E. coli*; *Bsu*, *Bacillus subtilis*; *6803*, *Synechocystis sp. PCC 6803*; *Sco*, *Streptomyces coelicolor*; *Pae*, *Pseudomonas aeruginosa*; *Tth*, *Thermus thermophilus*; *Taq*, *T. aquaticus*; *Cca*, *Cyanidium caldarium (chloroplaste)*; *Sim*, *Sinorhizobium meliloti*; *Atu*, *Agrobacterium tumefaciens*.

Ainsi, chez *E. coli*, il existe 7 facteurs σ qui permettent la reconnaissance de régions promoteur différentes (Helmann et Chamberlin, 1988; Lonetto *et al.*, 1992; Wosten, 1998; Paget et Helmann, 2003) (Fig. 4 et Tableau 1). Parmi ces 7 facteurs, 6 appartiennent à la famille σ^{70} et un seul représente la famille σ^{54} . La production de ces différents facteurs σ est soumise à des régulations, ce qui permet d'initier la transcription de certaines classes de gènes dans des conditions bien définies. Par ailleurs, un même gène peut être transcrit par différentes holoenzymes en fonction des conditions environnementales. Les caractéristiques de ces différents facteurs, ainsi que les conditions favorisant leur production dans la cellule, sont indiqués dans le tableau 1.

1.1.2.2.1.1) Le facteur σ^{70} chez *E. coli*.

Chez *E. coli*, le facteur σ^{70} permet l'initiation de la transcription de la majorité des gènes exprimés en phase exponentielle de croissance. Ce facteur est très conservé chez les bactéries (Lonetto *et al.*, 1992; Paget et Helmann, 2003). La structure 3D du facteur σ^{70} de *T. aquaticus* et d'*E. coli* a été déterminée par radiocristallographie (Campbell *et al.*, 2002; Malhotra *et al.*, 1996). Les deux structures sont très similaires et comportent quatre domaines structuraux reliés par trois chaînes flexibles (Fig. 5). Le domaine 1.1 n'est pas structuré et le domaine structural $\sigma 2$ comporte les régions conservées 1.2 et 2 (2.1-2.4) (Campbell *et al.*, 2002; Nudler et Gottesman, 2002). Le domaine structural $\sigma 3$ comporte les régions 3.0 et 3.1, tandis que la région 3.2 forme le domaine structural $\sigma 3.2$ (aussi appelé domaine de liaison). Le dernier domaine structural $\sigma 4$ est composé des régions 4.1 et 4.2. Le rôle de ces quatre régions conservées est indiqué sur la figure 5 et détaillé ci-dessous.

La région 1

Cette région comporte deux parties distinctes : la région 1.1 et la région 1.2.

La région 1.1 a une séquence faiblement conservée dans la famille σ^{70} . Elle n'est présente que dans les facteurs σ primaires et semble impliquée dans la modulation de la capacité de fixation à l'ADN du facteur σ^{70} libre (Dombroski *et al.*, 1992; Dombroski *et al.*, 1993). Il a été proposé qu'en interagissant avec la région 4.2 qui se fixe sur la boîte -35 du promoteur, la région 1.1 empêche la fixation aspécifique du facteur σ en absence des autres sous-unités (Dombroski *et al.*, 1993). D'après une étude par spectroscopie RMN et marquage isotopique

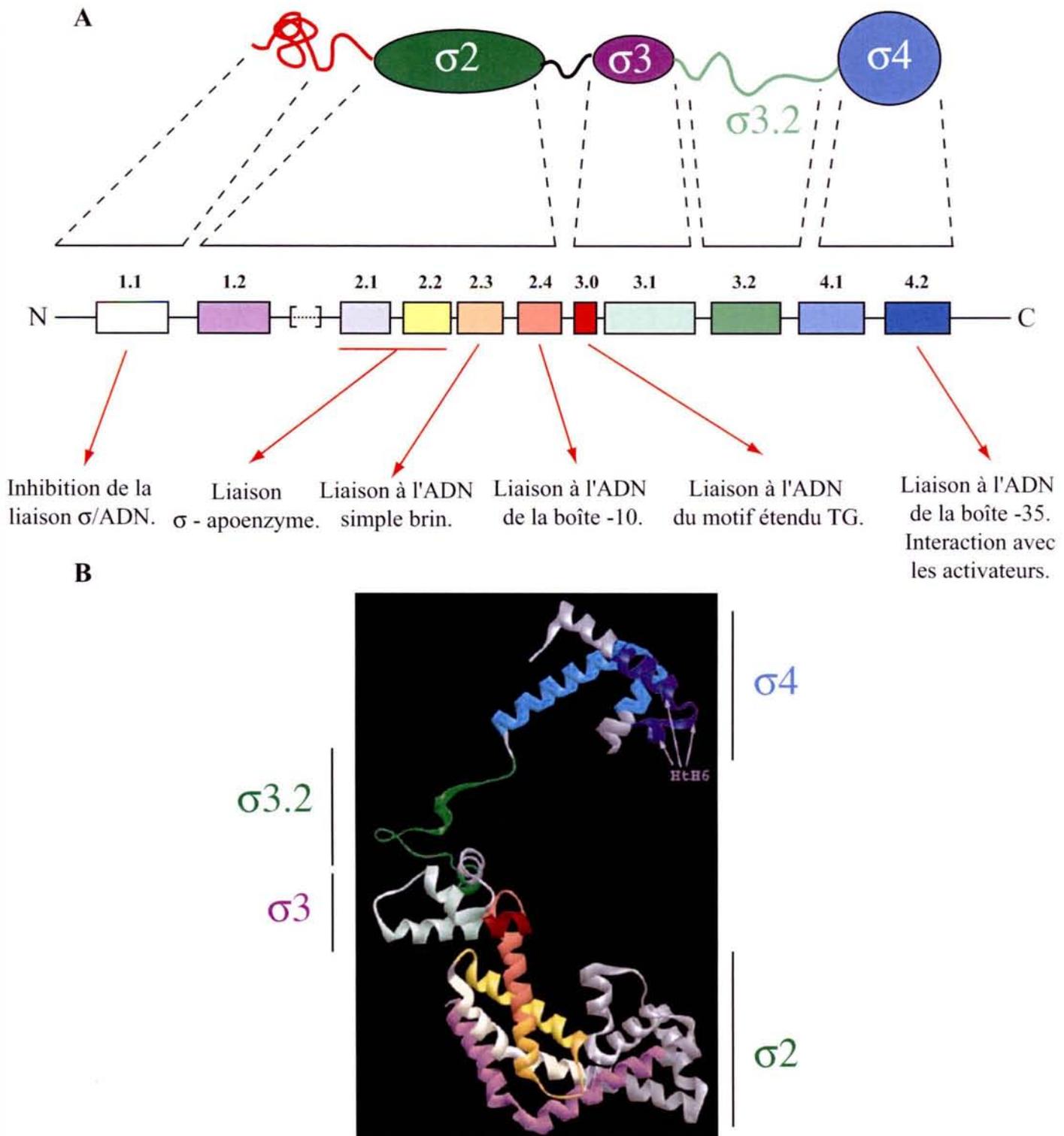


Figure 5 : Le facteur σ^{70} chez *E. coli*.

A) Organisation du facteur σ^{70} chez *E. coli* et rôle des différentes régions conservées dans les facteurs σ bactériens. L'organisation des régions conservées en domaines structuraux est représentée au-dessus de la séquence.

B) Structure du facteur σ^A de *T. thermophilus* (d'après Vassylyev *et al.*, 2002). Les couleurs sont identiques aux couleurs utilisées sur la représentation du panneau A. Les domaines structuraux sont indiqués ainsi que le motif HTH du domaine 4.2. Cette représentation est tirée de la structure de l'holoenzyme de *T. thermophilus* mais pas de la structure du facteur σ de *T. aquaticus* (Campbell *et al.*, 2002).

sur le facteur σ^{70} de *Thermotoga maritima*, cette interaction entre ces deux régions ne serait pas directe mais l'inhibition de la fixation aspécifique du facteur σ sur l'ADN par la région 1.1 serait indirecte par un effet d'encombrement stérique ou par un effet électrostatique (Camarero *et al.*, 2002). D'autres activités de la région 1.1 ont été mises en évidence par production de mutants (Wilson et Dombroski, 1997; Vuthoori *et al.*, 2001). La délétion d'une partie de la séquence inactive le facteur σ^{70} d'*E. coli in vivo*. Il en a été conclu que les acides aminés 50 à 100 sont importants pour l'initiation de la transcription notamment pour l'étape d'isomérisation du complexe ARN pol-promoteur en complexe RPo et aussi pour la formation du complexe d'élongation (Wilson et Dombroski, 1997). Pour d'autres promoteurs, la délétion de cette région augmente les capacités de formation du complexe RPo, ce qui suggère que le domaine 1.1 joue surtout un rôle en modulant l'interaction de l'holoenzyme avec le promoteur (Vuthoori *et al.*, 2001).

La région 1.2 est très conservée dans tous les facteurs de type σ^{70} mais aussi chez les bactéries. Les rôles de cette région seraient multiples. En effet, des mutations dans cette région peuvent affecter la liaison de l'holoenzyme au promoteur, la formation du complexe ouvert et le passage des complexes abortifs aux complexes d'élongation (Baldwin et Dombroski, 2001).

La région 2 :

La région 2 qui est fortement conservée chez les facteurs σ bactériens se décompose en 4 sous-régions nommées 2.1, 2.2, 2.3 et 2.4.

Les sous-domaines 2.1 et 2.2, notamment les résidus 361 à 390 sont impliqués dans la fixation du facteur σ à l'apoenzyme (Lesley et Burgess, 1989; Sharp *et al.*, 1999; Burgess et Anthony, 2001) mais aussi dans la structuration du domaine 2 (Malhotra *et al.*, 1996).

Plusieurs études ont montré que les régions 2.3 et 2.4 sont impliquées dans la reconnaissance de la boîte -10, ainsi que de régions d'ADN adjacentes, et aussi dans l'ouverture de la double hélice (pour revue : Record *et al.*, 1996). Certains acides aminés de ces régions sont impliqués à la fois dans des interactions avec l'ADN en double brin, avec l'ADN en simple brin et dans l'ouverture de la double hélice. Il a été proposé que la sous-région 2.4 est responsable de l'interaction avec les bases de la boîte -10 sous forme double brin. Cette interaction impliquerait les résidus Arg 441, Thr 440 et Gln 437 qui contacteraient les nucléotides aux positions -12 et -13 (Waldburger *et al.*, 1990; Marr et Roberts, 1997). D'autres acides aminés

semblent aussi impliqués dans la liaison de la partie de l'ADN double brin située en amont de la position -12. Il s'agit des résidus Arg 436, Trp 434 et Arg 451 (Fenton *et al.*, 2000).

Les acides aminés du domaine 2.3 sont surtout impliqués dans l'ouverture de la double hélice d'ADN notamment les acides aminés Tyr 425, Tyr 430, Trp 433 et Trp 434 (Juang et Helmann, 1994; Panaghie *et al.*, 2000; Fenton *et al.*, 2000). La juxtaposition de ces acides aminés aromatiques pourrait stabiliser l'ouverture de la double hélice d'ADN. En effet, la mutation de ces résidus diminue le taux de formation de complexe ouvert (RPo) (Panaghie *et al.*, 2000). L'acide aminé Tyr 425 serait aussi impliqué dans la liaison à l'ADN en double brin (Fenton *et al.*, 2000).

La région 3 :

La région 3 comporte le domaine 3.0 qui permet la reconnaissance du dinucléotide TG présent en amont de la boîte -10 caractéristique des promoteurs étendus (Barne *et al.*, 1997; Bown *et al.*, 1997). En utilisant un promoteur étendu dont le résidu G du motif TG avait été muté par un résidu C ou T et en réalisant une mutagenèse aléatoire du gène du facteur σ^{70} , Barne *et al.* (1997) ont montré que l'holoenzyme pouvait reconnaître le promoteur muté lorsque le facteur σ portait une mutation située à proximité de la région 2.4. Le segment correspondant est appelé 3.0 et les deux acides aminés responsables de la reconnaissance du motif TG sont Glu 458 et His 455. Les mêmes auteurs ont montré récemment par la méthode d'« alanine scanning » qu'en plus, les acides aminés Ile 439 et Arg 441 participent à la reconnaissance des paires de bases en position -14 et -15 (Sanderson *et al.*, 2003).

La région 3.1 conservée aurait un rôle structural et permettrait le bon positionnement de l'hélice contenant le domaine 3.0.

La région 3.2 forme un domaine de liaison entre les domaines $\sigma 2$ et $\sigma 4$ mais ne possède pas d'éléments caractéristiques de structure (Fig. 5) (Malhotra *et al.*, 1996; Vassylyev *et al.*, 2002; Campbell *et al.*, 2002). Lors de la fixation du facteur σ sur l'apoenzyme, la région 3.2 pénètre dans le site actif de l'holoenzyme et pourrait occuper le tunnel de sortie de l'ARN naissant (Vassylyev *et al.*, 2002; Murakami *et al.*, 2002a). L'ARN synthétisé se trouve ainsi bloqué dans le site actif, ce qui pourrait entraîner le phénomène de transcription abortive (qui sera décrit dans la partie mécanisme). De plus, cette région 3.2 pourrait se situer à proximité du nucléotide initiateur (Severinov *et al.*, 1994) et donc participer directement à l'initiation de la transcription (Borukhov et Severinov, 2002).

La région 4 :

La région 4 comporte le domaine 4.2 qui permet la reconnaissance de la boîte -35 (Keener et Nomura, 1993; Campbell *et al.*, 2002). Ce domaine forme une structure HTH (hélice tour hélice) que l'on trouve fréquemment dans des domaines de fixation à un ADN double brin (Fig. 5).

Le domaine 4.1 aurait un rôle dans la fixation du facteur σ à l'apoenzyme (Sharp *et al.*, 1999). En plus, cette région comporte un grand nombre de résidus hydrophobes qui permettent au domaine structural 4 d'adopter une forme incurvée (Murakami et Darst, 2003).

D'autre part, la région 4.2 semble impliquée dans les phénomènes d'activation transcriptionnelle (qui seront détaillés dans la suite de l'introduction). En effet, des mutations dans cette région affectent la liaison de l'holoenzyme à différents facteurs transcriptionnels (Li *et al.*, 1994; Kim *et al.*, 1995).

1.1.2.2.1.2) Les autres facteurs σ de la famille σ^{70} .

Le facteur σ^S est très similaire à σ^{70} en séquence et en structure et fait partie de la famille des facteurs primaires non essentiels (Fig. 4). La fonction des différentes sous-régions est a priori identique à celle du facteur σ^{70} . Les séquences des sous-régions impliquées dans la fixation à l'ADN sont très similaires, ce qui indique que les séquences des promoteurs sont très proches (Tableau 1).

Les facteurs σ^{32} et σ^{28} font partie de la famille des facteurs alternatifs (Fig. 4). Du point de vue structural, ces facteurs ne possèdent pas le domaine 1.1 (Gruber et Gross, 2003). L'absence de ce domaine régulateur indique que le mécanisme d'initiation de la transcription est différent pour ces facteurs. Les séquences reconnues par les holoenzymes associées à ces facteurs σ sont très différentes des séquences reconnues par $E\sigma^{70}$ (Tableau 1).

Les facteurs $\sigma^{24(E)}$ et $\sigma^{19(Fecl)}$ font partie de la famille des facteurs ECF. Les différences structurales majeures sont que ces facteurs ne possèdent ni région 1.1, ni domaine structural 3 (Gruber et Gross, 2003), ce qui implique des mécanismes d'initiation de la transcription différents pour ces holoenzymes comparés à $E\sigma^{70}$ notamment au niveau de la transcription abortive.

I.1.2.2.1.3) Le facteur σ^{54} chez *E. coli*.

La famille σ^{54} est très différente de la famille σ^{70} aussi bien en séquence qu'en structure (Wosten, 1998). Trois régions distinctes ont été mises en évidence au niveau des facteurs σ^{54} sur la base de la recherche d'homologie entre des facteurs σ d'espèces différentes. Ces régions, ainsi que leurs fonctions principales, sont indiquées dans la figure 6 (pour revue : Buck *et al.*, 2000).

La région I serait impliquée dans la liaison du facteur σ à l'apoenzyme, dans la liaison à l'ADN (Casaz et Buck, 1999) mais aussi dans la liaison aux activateurs de la transcription (Sasse-Dwight et Gralla, 1990; Syed et Gralla, 1998) et dans l'inhibition de l'étape d'isomérisation de l'holoenzyme (Cannon *et al.*, 1999). La région II est très variable et aurait un rôle dans l'ouverture de la double hélice (Wong et Gralla, 1992). La région III comporte trois motifs conservés : un motif HTH permettant la liaison à l'ADN, un motif X-link et une boîte RpoN. La boîte RpoN serait impliquée dans la reconnaissance des séquences permettant l'initiation de la transcription (Taylor *et al.*, 1996).

La séquence ADN située en amont du gène à transcrire, reconnue par le facteur σ^{54} associé à l'apoenzyme, est indiquée dans le tableau 1.

Certains gènes transcrits par l'holoenzyme $E\sigma^{54}$ chez *E. coli* sont impliqués dans le métabolisme de l'azote (Reitzer, 2003) mais d'autres gènes permettent la synthèse de protéines impliquées dans des fonctions très différentes comme une hydrogénase, des enzymes du catabolisme des acides gras ou des protéines de réponse aux phages (Reitzer et Schneider, 2001; Reitzer, 2003). La plupart de ces gènes non impliqués directement dans le métabolisme de l'azote sont impliqués dans des réponses à de multiples stress qui pourraient être liés à l'assimilation de l'azote (Reitzer et Schneider, 2001). Une autre possibilité serait que l'expression de ces gènes permette de réguler l'expression des gènes directement impliqués dans le métabolisme de l'azote (Reitzer, 2003).

I.1.2.2.2) L'holoenzyme $E\sigma^{70}$.

Comme il a été décrit précédemment, le facteur σ^{70} forme de nombreux contacts avec l'apoenzyme (Owens *et al.*, 1998; Sharp *et al.*, 1999; Burgess et Anthony, 2001; Gruber *et al.*, 2001). Les mutations obtenues par mutagenèse aléatoire et sélectionnées pour leur capacité à affecter la liaison du facteur σ^{70} sur l'apoenzyme se sont avérées être localisées tout le long de

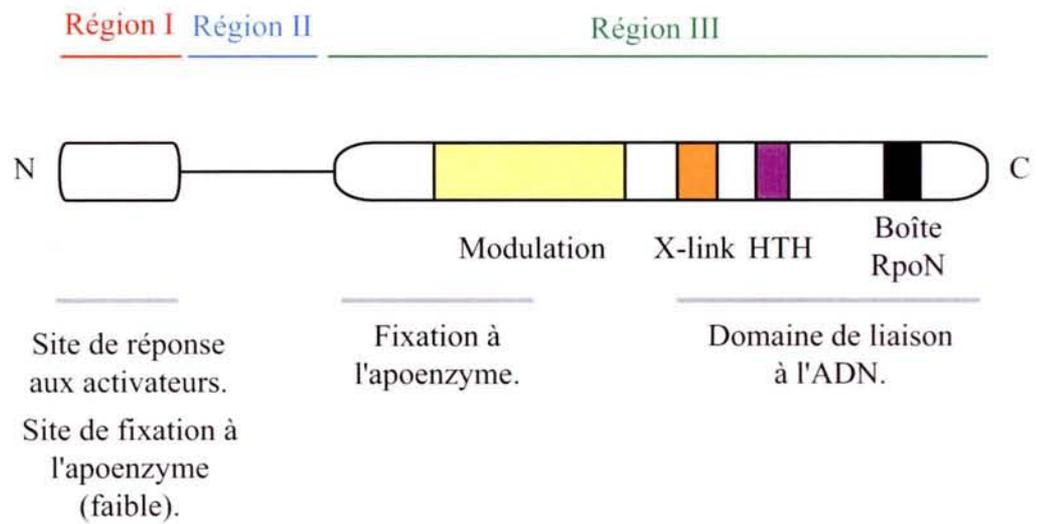


Figure 6 : Le facteur σ^{54} chez *E. coli*.

Organisation du facteur σ^{54} chez *E. coli* et rôle des différentes régions conservées dans les facteurs σ^{54} bactériens.

la séquence de ce facteur, en particulier, au niveau des régions 2.1, 2.2, 2.4, 3.1, 4.1 et 4.2 (Sharp *et al.*, 1999). Le contact entre les régions 2.1 et 2.2 et la séquence β' 260-309 (Région conservée B) semble être le site majeur d'interaction entre les 2 partenaires (Burgess et Anthony, 2001). Gruber *et al.*, (2001) ont utilisé une méthode d'immunoprécipitation pour identifier les domaines de σ^{70} en liaison avec l'apoenzyme. Cette méthode a permis de mettre en évidence d'autres interactions notamment entre la région 1.1 du facteur σ^{70} et les domaines E, F et G de la sous-unité β .

La conclusion de ces études est que les interactions entre le facteur σ^{70} et l'apoenzyme sont très nombreuses et réparties tout au long de la séquence du facteur σ .

Deux méthodes différentes ont été utilisées pour déterminer la structure de l'holoenzyme associée au facteur σ^{70} (pour revue : Young *et al.*, 2002; Murakami et Darst, 2003; Borukhov et Nudler, 2003). D'une part, la structure globale de l'holoenzyme $E\sigma^{70}$ a pu être étudiée par cryomicroscopie électronique suivie d'une reconstitution angulaire et un modèle à 9,5 Å a pu être obtenu (Finn *et al.*, 2000). Cette résolution ne permet pas une identification des chaînes latérales des différents résidus. Les auteurs concluent que le passage de la forme apoenzyme à la forme holoenzyme (c'est-à-dire l'association avec σ^{70}) entraîne un changement de conformation de la sous-unité β' , alors que le positionnement relatif des autres sous-unités change peu (Finn *et al.*, 2000).

La structure de l'holoenzyme $E\sigma^A$ de *T. thermophilus* a été étudiée à très haute résolution par radiocristallographie (2,6 Å) (Vassylyev *et al.*, 2002) et celle de *T. aquaticus* l'a été à une résolution un peu plus faible (4Å) (Murakami *et al.*, 2002a). Ces résolutions permettent le positionnement d'un certain nombre de chaînes latérales ainsi que de la chaîne principale.

La structure de la région 1.1 de l'holoenzyme de *T. thermophilus* n'a pas pu être résolue et cette région était absente de l'holoenzyme de *T. aquaticus* étudiée en radiocristallographie.

D'après les structures obtenues pour ces deux holoenzymes, les différents domaines structuraux du facteur σ ($\sigma 1$ à $\sigma 4$) se situent sur une seule face de l'apoenzyme (Fig. 7). Ces structures montrent que lors du passage apoenzyme-holoenzyme, des glissements et des rotations des domaines les uns par rapport aux autres se produisent, ce qui pourrait conduire à une fermeture du canal permettant la fixation de l'ADN (Murakami *et al.*, 2002a; Hsu, 2002a; Vassylyev *et al.*, 2002). Ceci pourrait permettre de réduire la fixation non spécifique à l'ADN. Les domaines mobiles dans l'apoenzyme, décrits ci-dessus possèdent toujours des structures indépendantes dans l'holoenzyme et s'associent avec les domaines du facteur σ (Murakami et

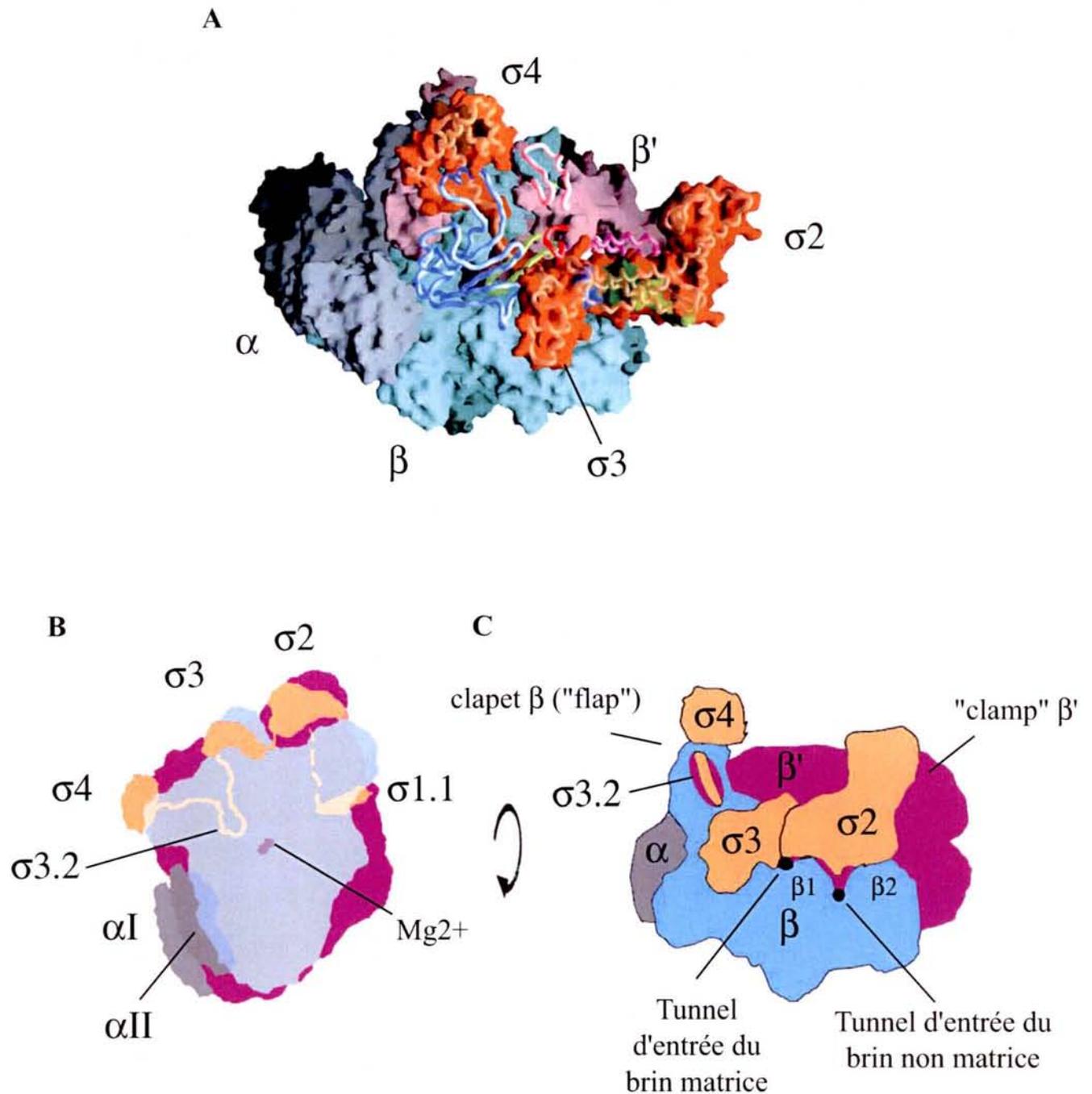


Figure 7 : Structure de l'ARN pol sous sa forme holoenzyme (d'après Murakami *et al.*, 2002a et Young *et al.*, 2002).

A) Structure de l'holoenzyme de *T. aquaticus* (Murakami *et al.*, 2002a). Les différentes sous-unités sont indiquées en couleurs : gris α , bleu β , violet β' et orange σ .

B) Représentation de l'holoenzyme. La sous-unité β en bleu a été rendue transparente pour visualiser le site actif. Les mêmes couleurs que pour A) sont utilisées (d'après Young *et al.*, 2002).

C) Représentation de l'holoenzyme avec une rotation de 90° par rapport à B. Les couleurs utilisées sont les mêmes que pour A) (d'après Hsu, 2002).

Darst, 2003). Par exemple, le domaine $\sigma 2$ se fixe sur une structure appelée « clamp » formée par β' et $\sigma 3$ interagit avec la sous-unité β (domaine $\beta 1$). De plus, le clapet β s'associe avec le domaine $\sigma 4$. L'ensemble ainsi formé est mobile, ce qui permet d'accomoder les séquences des promoteurs (Murakami et Darst, 2003).

Il est à noter que dans la structure des holoenzymes, le deuxième ion Mg^{2+} essentiel à la catalyse n'est pas visible ou est mal positionné pour la catalyse. En effet, l'affinité du site de fixation pour cet ion serait faible. D'autres études par radiocristallographie ont confirmé le positionnement de ces deux ions catalytiques (Artsimovitch *et al.*, 2004).

La région $\sigma 3.2$ qui n'est pas structurée, pénètre dans le site actif de l'ARN polymérase et se situe à proximité de l'ion catalytique Mg^{2+} (Fig. 7 B). Cette région pourrait avoir une influence sur la fixation du nucléotide initiateur (Severinov *et al.*, 1994; Murakami *et al.*, 2002a). Des études ont confirmé des proximités spatiales entre le nucléotide initiateur et σ (Severinov *et al.*, 1994). De plus, cette région 3.2 semble fermer le canal permettant la sortie de la molécule d'ARN (Fig. 7 C), ce qui entraînerait le phénomène de transcription abortive (Murakami *et al.*, 2002a).

Lors du passage de la forme apoenzyme à la forme holoenzyme, le changement structural le plus important pour le mécanisme d'initiation de la transcription est celui rendant accessible les domaines de liaison à l'ADN du facteur σ . Ces régions (2.4 et 4.2) vont en effet pouvoir interagir avec le promoteur notamment avec les régions -10 et -35.

La technique de FRET (Fluorescence Resonance Energy Transfer) a été utilisée pour identifier la position du facteur σ^{70} sur l'apoenzyme de *T. aquaticus* (Mekler *et al.*, 2002). Cette technique est basée sur la mesure des distances entre deux fluorophores, l'un étant donneur, l'autre accepteur. Pour cela, la longueur d'onde d'émission du donneur doit recouvrir la longueur d'onde d'excitation de l'accepteur. Dans ces conditions, il peut y avoir un transfert d'énergie entre le donneur et l'accepteur. La mesure de l'efficacité du transfert permet d'estimer la distance entre les deux fluorophores. Les fluorophores ont été placés à différents endroits de l'apoenzyme et du facteur σ^{70} . Après avoir calculé les distances entre les différentes positions des fluorophores, les auteurs ont replacé (par docking informatique) le facteur σ^{70} sur la structure de l'apoenzyme de *T. aquaticus* (Zhang *et al.*, 1999). L'avantage de cette technique par rapport aux approches cristallographiques est qu'elle permet d'obtenir des informations sur la structure en solution. L'approche a aussi permis de définir la position de la région 1.1 très mobile (non observée par radiocristallographie). Cette région très acide

est localisée au niveau de l'entrée du canal de fixation de l'ADN (Mekler *et al.*, 2002). Elle pourrait ainsi bloquer la fixation non spécifique de l'holoenzyme sur l'ADN.

En conclusion, la structure des holoenzymes révèle que les domaines de fixation non spécifique à l'ADN sont protégés, alors que les domaines de fixation spécifiques à l'ADN sont exposés au solvant, ce qui permet la reconnaissance très spécifique des régions promoteurs.

I.1.2.2.3) Les autres holoenzymes.

Sachant que les différences entre les facteurs σ semblent se situer surtout au niveau d'éléments de régulation de leur activité (domaine 3 et sous-région 1.1), la structure globale des holoenzymes de la famille σ^{70} semble être conservée. En effet, dans tous les facteurs σ de la famille σ^{70} chez *E. coli*, les domaines $\sigma 2$ et $\sigma 4$ sont retrouvés, ce qui impliquerait une reconnaissance structurale de l'apoenzyme très similaire à celle du facteur σ^{70} . Les structures des différentes holoenzymes associées à des facteurs de la famille σ^{70} seraient donc plus ou moins similaires, ainsi que le mécanisme de synthèse de l'ARN (Gruber et Gross, 2003). Les différences ne seraient dues qu'aux particularités des régions de fixation à l'ADN permettant la reconnaissance de différents promoteurs et qu'à des différences de mécanisme de la régulation de l'initiation de la transcription.

Par exemple, pour l'holoenzyme $E\sigma^S$, les différences avec $E\sigma^{70}$ semblent limitées aux acides aminés qui permettent la reconnaissance des boîtes -10 et -35 et à ceux qui régulent les étapes de l'initiation de la transcription (Lee et Gralla, 2001; Lee et Gralla, 2002; Checroun *et al.*, 2004). Pour la famille $E\sigma^{54}$, il semble que la structure globale de l'holoenzyme soit tout à fait différente en dépit de l'identité de l'apoenzyme.

I.2) Le mécanisme d'initiation de la transcription.

Comme il vient d'être décrit, deux grands types d'holoenzymes ont été définis, les holoenzymes $E\sigma^{54}$ et $E\sigma^{70}$, qui diffèrent par leur structure mais aussi par leur mécanisme d'initiation de la transcription.

L'initiation de la transcription est un mécanisme complexe qui met en œuvre plusieurs complexes ARN polymérase/ADN successifs qui aboutissent au déplacement de l'holoenzyme le long de l'ADN et à la formation de l'ARNm (Fig. 8).

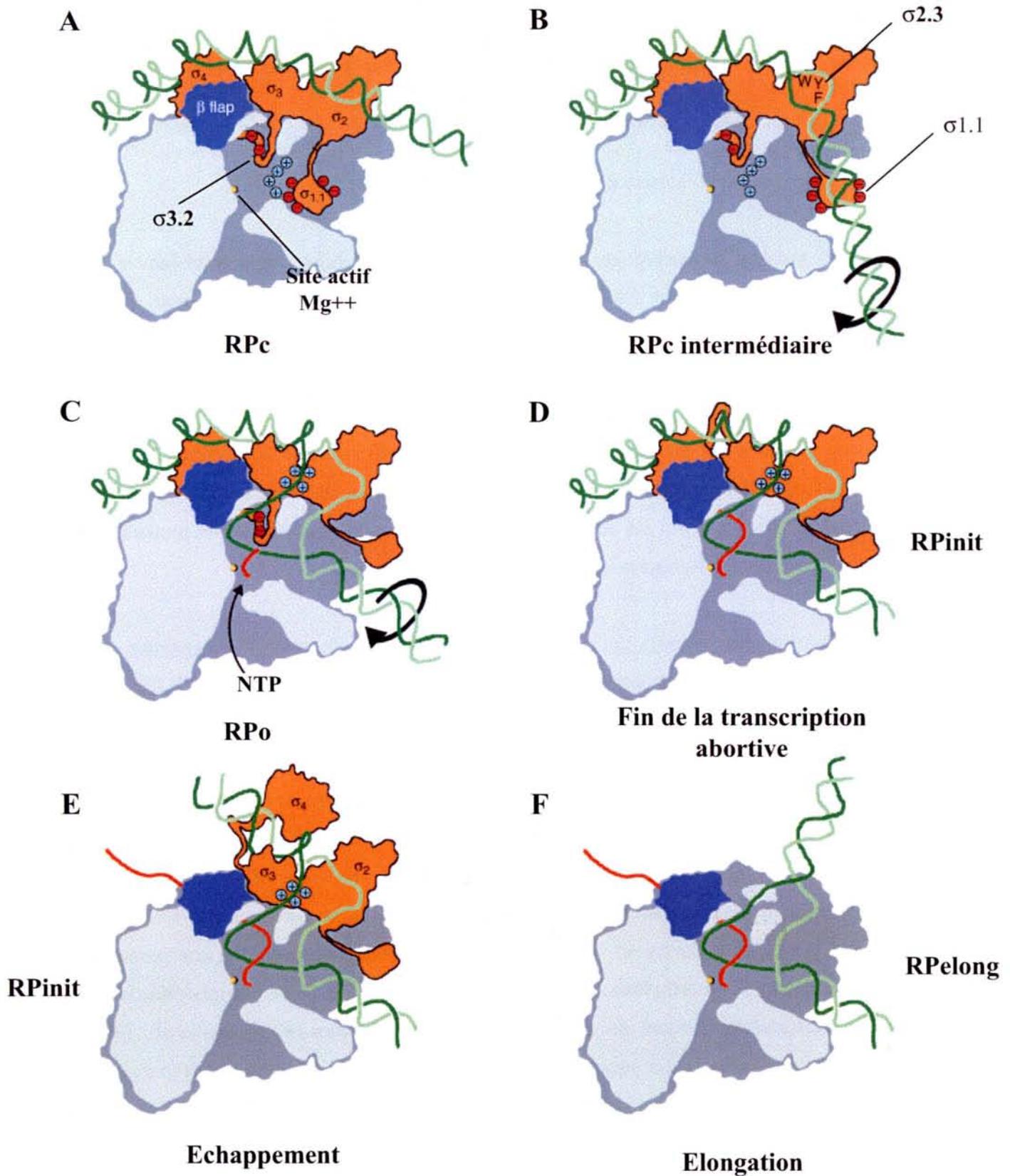


Figure 8 : Modèle du mécanisme d'initiation de la transcription chez les bactéries (d'après Murakami et Darst, 2003).

Les différentes étapes de ce mécanisme sont indiquées par des lettres. L'apoenzyme est représentée en gris. Cette représentation ne montre pas la sous unité β qui se trouve au premier plan et cacherait donc le site actif. Le facteur σ est représenté en orange et le clapet β ("flap") en bleu. Le brin matrice de l'ADN est en vert foncé et le brin non matrice en vert clair. La molécule d'ARN en cours de synthèse est représentée en rouge.

L'holoenzyme $E\sigma^{70}$, ainsi que les holoenzymes associées aux facteurs σ primaires ont été très étudiées.

I.2.1) Mécanisme d'action de l'holoenzyme $E\sigma^{70}$.

Un grand nombre de techniques ont été utilisées pour aboutir à une vision la plus cinématographique possible des mécanismes moléculaires impliqués dans cette réaction complexe.

1) Etudes biochimiques.

Beaucoup d'études sur les propriétés de mutants de l'ARN pol et de diverses séquences promoteurs ont été réalisées. De plus, divers analogues de nucléotides ont été utilisés. Des techniques d'empreintes ont été aussi utilisées pour définir les régions d'ADN protégées par l'enzyme (pour revue : Record *et al.*, 1996).

2) Modélisation des données de radiocristallographie obtenues pour l'holoenzyme de *T. aquaticus* associée à un fragment d'ADN ayant une architecture en fourche (Fig. 9 A).

Ce fragment d'ADN correspond à un promoteur comprenant une boîte -10 étendue et une boîte -35 de séquence canonique dont le segment allant des positions -41 à -12 est en double brin et celui allant des positions -11 à -7 en simple brin (Murakami *et al.*, 2002b). Cet ADN en fourche mime l'ADN dans le complexe ouvert dont une partie de la boîte -10 est ouverte. L'ARN pol a une forte affinité pour ce type de molécules et le complexe ADN-ARN pol formé possède les caractéristiques du complexe ouvert (Guo et Gralla, 1998; Helmann et deHaseth, 1999; Tsujikawa *et al.*, 2002). Après étude à résolution moyenne par radiocristallographie du complexe ADN en fourche-ARN pol (Fig. 9 B), une modélisation des données a été réalisée incluant des résultats d'études biochimiques (Murakami *et al.*, 2002b) (Fig. 9). Cette modélisation a permis de mettre en évidence les éléments importants pour le mécanisme d'initiation de la transcription (Murakami *et al.*, 2002b; Murakami et Darst, 2003).

3) Technique de FRET (Fluorescence Resonance Energy Transfer).

Cette technique a permis d'obtenir des données sur le positionnement de la molécule d'ADN dans l'holoenzyme en solution (Mekler *et al.*, 2002). Les fluorophores ont été placés à différents endroits de l'holoenzyme et de l'ADN. Ces études ont mis en évidence des

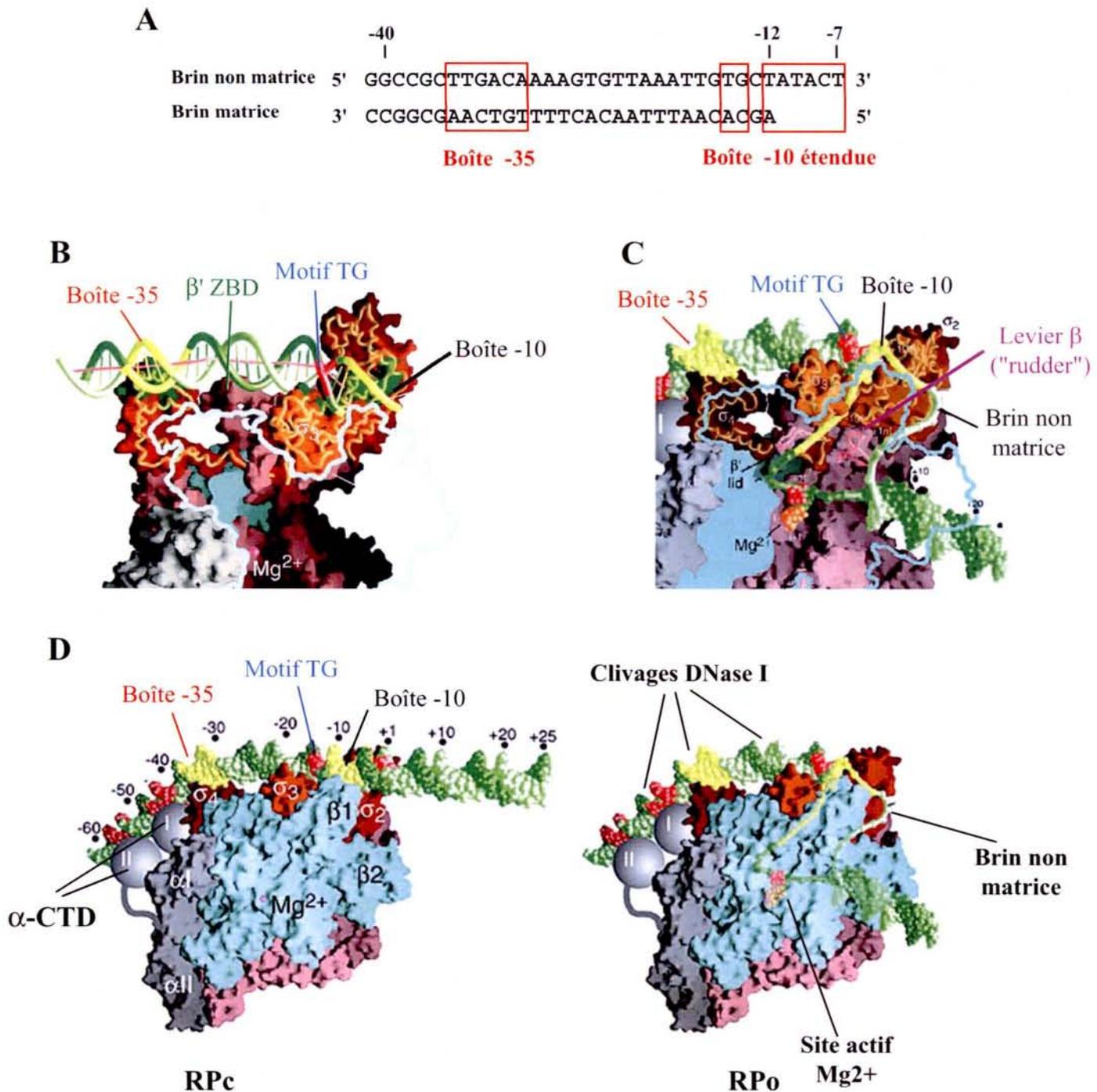


Figure 9 : Structure et modèle de l'holoenzyme isolée de *T. aquaticus* en association avec un ADN en fourche (d'après Murakami *et al.*, 2002b).

A) Séquence de l'ADN en fourche utilisé pour obtenir la structure du complexe RPO.

B) Structure de l'ARN pol associée à la molécule d'ADN en fourche. Les sous-unités β' , α et σ sont représentées en rose, gris et orange respectivement. Le contour de la sous-unité β est représenté en bleu. Le brin matrice est représenté en vert foncé et le brin non matrice en vert clair. Les séquences des boîtes -10 et -35 sont indiquées en jaune et le motif étendu TG est indiqué en rouge.

C) Modèle du complexe RPO montrant la bulle de transcription. Les deux brins d'ADN sont séparés physiquement dans deux tunnels différents de l'ARN pol. Les couleurs utilisées sont les mêmes que pour B). Les différents éléments discutés dans le texte sont représentés.

D) Modèle récapitulant le passage du complexe RPO au complexe RPo. L'ADN est représenté des positions -60 à +25. Les couleurs utilisées sont les mêmes que pour B) sauf que la sous-unité β n'est pas transparente. Le site d'initiation de la transcription ainsi que l'élément UP qui permet la fixation des sous-unités α -CTD sont représentés en rouge. La disposition possible de ces sous-unités (I et II) sur l'ADN en amont de la boîte -35 est indiquée. Les coupures à la DNaseI, qui apparaissent au cours du mécanisme d'initiation de la transcription aux positions -25, -35 et -45, sont indiquées.

mouvements des zones flexibles de l'holoenzyme au cours du mécanisme d'initiation de la transcription.

4) Techniques de pontages covalents à des sites spécifiques.

Ces techniques qui permettent le pontage covalent entre une sonde fixée à un site défini sur la molécule d'ADN (ou ARN) et l'ARN pol ont été utilisées pour étudier les interactions entre l'ADN et les différents éléments de l'ARN pol au sein du complexe RPo (Naryshkin *et al.*, 2000) et les interactions ADN, ARN naissant et ARN pol au sein du complexe d'élongation (Korzheva *et al.*, 2000).

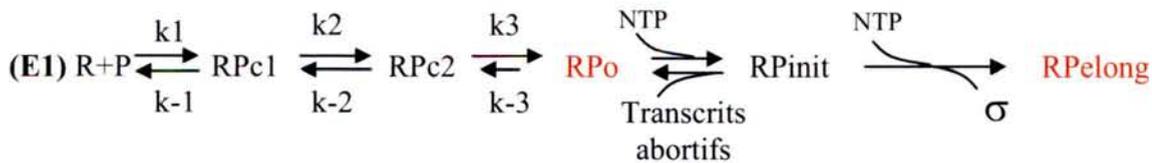
Toutes ces techniques ont apporté des éléments d'information sur les structures des différents complexes observés au cours du mécanisme d'initiation de la transcription.

5) Etudes cinétiques.

L'étude de la cinétique de formation des différents complexes repose largement sur l'utilisation de techniques de modifications par des agents chimiques et coupures enzymatiques, ainsi que des techniques de protection contre ces modifications. Certaines de ces études ont été réalisées au cours de cinétiques rapides. D'autres études ont été réalisées avec des mutants de l'ARN pol ou des promoteurs différents (pour revue : Record *et al.*, 1996). La réalisation de telles expériences à différentes températures et dans différents tampons a permis la mise en évidence de complexes formés très rapidement et de complexes à très courte durée de vie (Craig *et al.*, 1998; Buckle *et al.*, 1999; Saecker *et al.*, 2002).

L'utilisation de pontage aux UV a permis de travailler en cinétique très rapide et ainsi de mettre en évidence les différents intermédiaires se formant successivement lors de la réaction d'initiation de la transcription (Buckle *et al.*, 1991a; Buckle *et al.*, 1991b; Buckle *et al.*, 1999). Cette technique repose sur l'emploi d'un flash laser UV (266 nm) de forte intensité, qui permet de créer des liaisons covalentes au sein de l'ADN (formation de dimères de thymine), mais aussi entre deux molécules (protéine-ADN) assemblées sous forme de complexe. Le pontage ne peut se produire que si des acides aminés et des bases de l'ADN sont proches dans l'espace (Buckle *et al.*, 1999). Cette technique permet donc la mise en évidence de contacts au sein des complexes promoteur-ARN pol. De plus, son couplage avec des systèmes de mélange rapide a permis de réaliser des études cinétiques de formation de ces complexes. D'autres agents tels que le formaldéhyde permettant le pontage entre promoteur et l'ARN pol ont aussi été utilisés (Brodolin et Buckle, 2001).

Ces données cinétiques ont permis de définir les différentes étapes de l'initiation de la transcription et de les représenter par l'ensemble de réactions suivant (E1) (Buc et McClure, 1985; Record *et al.*, 1996; Craig *et al.*, 1998; Buckle *et al.*, 1999; Brodolin et Buckle, 2001; Saecker *et al.*, 2002) :



(avec R = RNA polymérase, P = Promoteur, RPc = complexe fermé (closed), RPo = complexe ouvert (open), RPinit = complexe d'initiation et RPelong = complexe d'élongation)

Cet ensemble de réactions peut se décomposer en deux phases :

1) Tout d'abord, l'ouverture de la double hélice (formation du complexe RPo) rassemble les étapes de liaison de l'ARN pol sur l'ADN (reconnaissance du promoteur), de changement de conformation de l'holoenzyme (isomérisation), ainsi que de changement de conformation de l'ADN. En effet, au cours de ce mécanisme, l'ADN est courbé par l'ARN pol, ce qui entraînerait son enroulement autour de l'holoenzyme (Rivetti *et al.*, 1999; Coulombe et Burton, 1999). La présence de cet enroulement, observé par microscopie électronique (Rivetti *et al.*, 1999), faciliterait l'ouverture de la double hélice (Coulombe et Burton, 1999). L'étape de liaison à l'ADN est rapide, alors que le passage au complexe RPo est une étape lente. La stabilité des complexes RPo est en général élevée.

2) Ensuite, le passage en élongation rassemble les étapes de formation des transcrits abortifs (RPinit) et d'échappement du promoteur (RPelong).

Les barrières d'énergie correspondant à ces deux grandes transitions sont assez élevées, ce qui indique que certaines des étapes impliquées peuvent être limitantes et fortement régulées. Pour un grand nombre de promoteurs, l'étape limitante est la formation du complexe RPo. Pour d'autres, le passage à l'étape d'élongation est l'étape limitante.

I.2.1.1) Processus de formation des complexes RPo.

Lors de ce processus de formation des complexes RPo, l'holoenzyme reconnaît spécifiquement les séquences du promoteur. Nous allons tout d'abord décrire les acides

aminés des différentes sous-unités de l'ARN pol importants pour la reconnaissance de ces séquences.

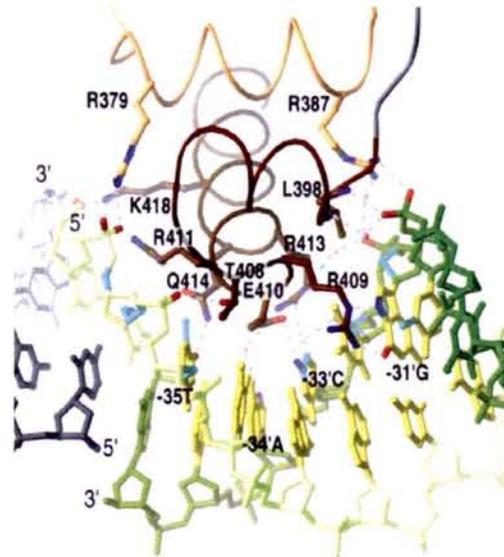
Différents intermédiaires ont été mis en évidence dans le processus de formation des complexes ouverts RPo. Les propriétés de ces intermédiaires ont été déterminées à basse température pour augmenter leur durée de vie mais ils pourraient ne pas être représentatifs des intermédiaires qui existent à 37°C. De plus, les propriétés de ces intermédiaires pourraient être différentes d'un promoteur à l'autre suivant la séquence des boîtes -10 et -35. Dans un deuxième temps, nous décrirons les connaissances sur ces complexes déterminés à basse température et nous présenterons des résultats cinétiques obtenus à 37°C.

Comme il a déjà été indiqué, la liaison du facteur σ^{70} à l'apoenzyme entraîne un changement de sa conformation rendant accessibles les acides aminés des régions 2.4 et 4.2, nécessaires à la reconnaissance de l'ADN. Ainsi, ces acides aminés vont pouvoir interagir avec les boîtes -10 et -35 (Burgess et Anthony, 2001).

Reconnaissance de la boîte -35.

L'existence d'un contact entre la région 4.2 du facteur σ^{70} et la boîte -35 supposé être mis en place très rapidement au cours de la réaction de formation des complexes RPo (Eichenberger *et al.*, 1997; Buckle *et al.*, 1999) a été vérifiée par des expériences de mutagenèse compensatoire entre σ^{70} et l'ADN (Siegele *et al.*, 1989; Gardella *et al.*, 1989) et par pontage aux UV (technique laser) (Buckle *et al.*, 1999). Enfin, récemment, un fragment d'ADN a été cristallisé en présence du domaine $\sigma 4$ du facteur σ de *T. aquaticus* (Campbell *et al.*, 2002) (Fig. 10). Les interactions ADN-protéine se situent surtout au niveau du grand sillon de l'ADN et concernent l'ensemble de la boîte -35. Il a ainsi pu être montré que l'Arg409 contacte le résidu G-31 situé sur le brin matrice (Fig. 10), ce qui est en accord avec les données de protection (Siegele *et al.*, 1989) et aussi avec le fait que la méthylation du résidu G-31 en position N7 par le DMS empêche aussi la fixation de l'holoenzyme (Siebenlist *et al.*, 1980). Le domaine $\sigma 4.2$ contacte aussi la chaîne phosphodiester sur le brin non matrice (Campbell *et al.*, 2002). La reconnaissance de la boîte -35 par la région 4.2 du facteur σ induit la formation d'une courbure de 36° de l'hélice d'ADN qui faciliterait l'enroulement de l'ADN autour de l'ARN pol.

A



B

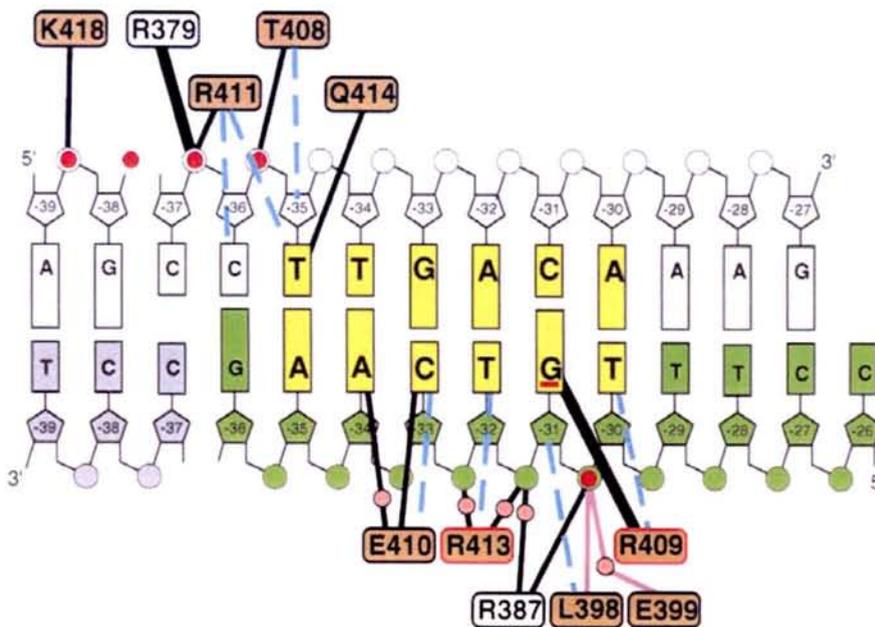


Figure 10 : Interaction entre le domaine σ_4 du facteur σ^A et l'ADN de la boîte -35 d'un promoteur (d'après Campbell *et al.*, 2002).

A) Structure du complexe σ_4 de *T. aquaticus* et une région d'ADN contenant une boîte -35 (Campbell *et al.*, 2002). La chaîne principale du facteur σ_4 est représentée en brun et doré. Le brin non matrice est représenté en vert clair, le brin matrice en vert foncé et les bases de la boîte -35 en jaune

B) Représentation schématique des interactions entre le domaine σ_4 et l'ADN de la boîte -35. Les contacts se situent dans le sillon majeur de l'ADN. Le brin non matrice est représenté en blanc, le brin matrice en vert et les bases de la boîte -35 en jaune. Les bases en gris ont été ajoutées par symétrie. Les résidus des régions 4.1 et 4.2 qui interagissent avec l'ADN sont indiqués par des étiquettes blanches et brunes. La numérotation est celle du facteur σ de *T. aquaticus*. Le résidu Arg 409 pour *Taq* correspond au résidu 584 chez *E. coli* (R379/R554, R387/R562, L398/L573, E399/E574, T408/T583, R409/R584, E410/E585, R411/R586, R413/R588, Q414/Q589, et K418/K593). Les molécules d'eau sont indiquées par des points roses et les interactions par des lignes épaisses.

Reconnaissance de la boîte -10.

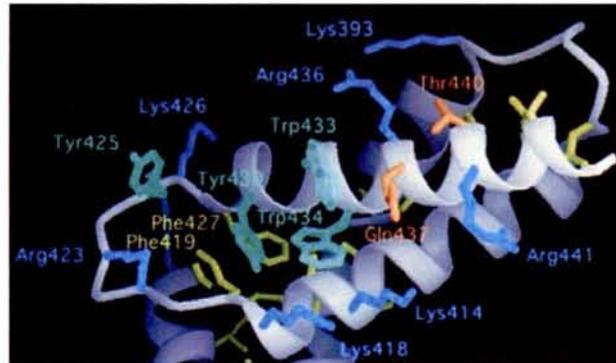
La boîte -10 du promoteur serait reconnue sous forme double brin par le domaine 2.4 du facteur σ^{70} . Ensuite, des interactions sont formées avec l'ADN de la boîte -10 sous forme simple brin lors du passage au complexe RPo. Les déterminants de séquence de la boîte -10 qui influencent fortement la fixation à l'ADN double brin lors des étapes initiales de reconnaissance sont les positions -12 et -11 (TATAat) (Helmann et deHaseth, 1999; Fenton et Gralla, 2001).

Les acides aminés impliqués dans la reconnaissance de la boîte -10 sous sa forme double brin sont principalement Arg 436, Gln 437, Trp 434, Thr 440, Arg 441 et Arg 451 du domaine $\sigma^{2.4}$ (Waldburger *et al.*, 1990; Marr et Roberts, 1997; Fenton *et al.*, 2000). La structure 3D du domaine σ^2 d'*E. coli* obtenue par radiocristallographie montre que les acides aminés qui permettent la reconnaissance de la boîte -10 sont situés sur une même face d'une hélice α (Fig. 11). Un modèle d'interaction entre $\sigma^{2.4}$ et la boîte -10 a été construit à partir de ces données (Malhotra *et al.*, 1996) (Fig. 11). Ce modèle est en accord avec la structure obtenue pour l'ARN pol associée à un ADN en fourche. Dans ces structures, les acides aminés Gln 437 et Thr 440 de la région $\sigma^{2.4}$ sont positionnés à proximité du sillon majeur au niveau de la position -12 (Murakami *et al.*, 2002b). Néanmoins, cette proximité ne semble pas refléter l'existence d'une interaction importante pour la fixation à l'ADN double brin puisque la mutation de ces acides aminés a peu d'effet sur cette fixation (Fenton *et al.*, 2000).

Reconnaissance du dinucléotide TG.

Dans le cas des promoteurs étendus, le dinucléotide TG pourrait constituer un autre point d'encrage pour la reconnaissance par l'ARN polymérase et ainsi augmenterait l'affinité de l'holoenzyme pour le promoteur. Grâce à ces contacts supplémentaires, la présence d'une boîte -35 ne serait pas indispensable à l'activité des promoteurs étendus (Bown *et al.*, 1997). La figure 12 montre le positionnement des acides aminés His 455, Glu 458, Ile 439 et Arg 441 (cités précédemment) et du dinucléotide TG dans la structure de l'holoenzyme de *T. aquaticus* associée avec un fragment d'ADN en fourche (Sanderson *et al.*, 2003). Les acides aminés His 455 et Glu 458 sont exposés à la surface d'une hélice α et font face au grand sillon de l'ADN au niveau du motif étendu. De plus, l'acide aminé Glu 458 pourrait se situer à proximité du résidu T en position -15 (Murakami *et al.*, 2002b).

A



B

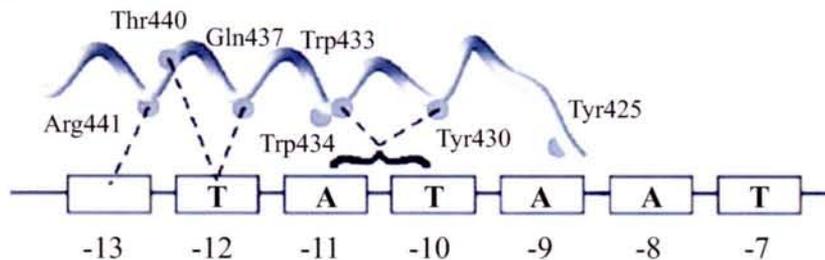


Figure 11 : Interaction entre le domaine $\sigma 2$ du facteur σ^{70} et l'ADN de la boîte -10 (d'après Malhotra *et al.*, 1996).

A) Structure du domaine $\sigma 2$ du facteur σ d' *E. coli* (Malhotra *et al.*, 1996). Les résidus en jaune forment le coeur de l'hélice hydrophobe. Les résidus en vert sont les résidus aromatiques conservés de la région 2.3 importants pour l'ouverture de la double hélice et les résidus qui pourraient interagir avec la base en position -12 de la boîte -10 sont indiqués en orange. Les résidus basiques en bleu pourraient être impliqués dans les interactions avec la chaîne phosphodiester de l'ADN.

B) Représentation schématique des interactions potentielles entre le domaine $\sigma 2$ et l'ADN de la boîte -10. A partir de la structure en A), les auteurs ont positionné les bases du brin non matrice de la boîte -10 par rapport aux hélices formées par les régions conservées de $\sigma 2$. Les interactions qui ont été déterminées par des approches biochimiques ou génétiques sont indiquées par des lignes pointillées.

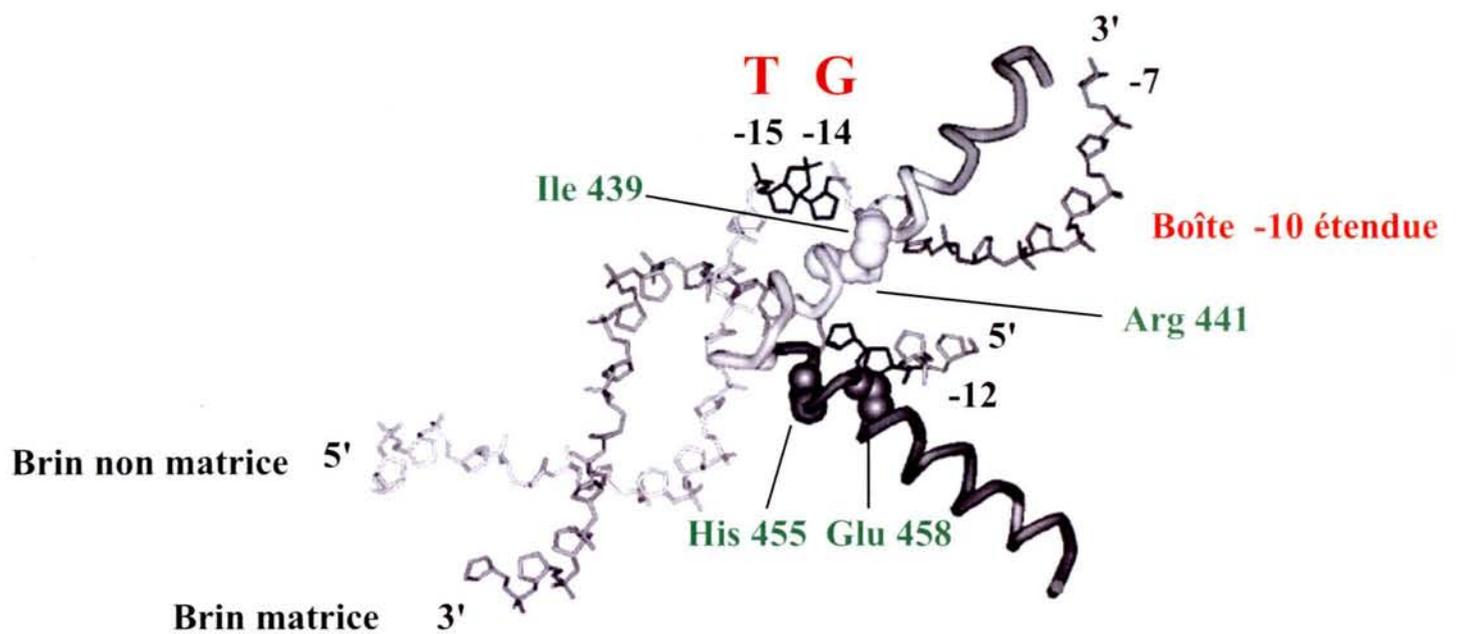


Figure 12 : Interactions entre le domaine $\sigma^{3.0}$ du facteur σ^A de *T. aquaticus* et le motif TG (d'après Sanderson *et al.*, 2003).

Le brin matrice est représenté en gris foncé et le brin non matrice en gris clair. Les nucléotides T -14 et G -15 sont représentés en noir. Les acides aminés qui semblent avoir un rôle dans la reconnaissance de l'élément TG sont indiqués.

Reconnaissance d'éléments présents en amont de la boîte -35.

Des contacts additionnels entre l'ARN pol et la séquence située en amont de l'élément -35 du promoteur sont médiés par le domaine C-terminal des sous-unités α (Fig. 9 D). Ces contacts, qui peuvent être spécifiques (avec l'élément UP) ou non, induisent la formation d'une courbure de la molécule d'ADN qui pourrait faciliter l'enroulement de la molécule d'ADN autour de l'ARN pol.

Ces interactions qui vont permettre la reconnaissance spécifique du promoteur et l'ouverture de la double hélice se produisent séquentiellement au cours du mécanisme de formation du complexe ouvert RPo. Lors de la formation du complexe RPo, l'ADN est courbé autour de l'ARN pol mais les étapes précises où apparaissent ces courbures ne sont pas définies. Il a été proposé que les premiers contacts ARN pol/promoteur s'établissent entre la boîte -35 et le domaine $\sigma 4.2$ de l'holoenzyme (pour revue : deHaseth *et al.*, 1998). Les contacts entre l'ARN pol et la boîte -10 sous forme double brin pourraient s'établir au même moment. Ces interactions vont permettre la formation de différents complexes qui, dans un premier temps, ont été caractérisés à basse température. La première des réactions consiste en la liaison de l'holoenzyme au promoteur pour former le complexe fermé (RPc). Dans ce complexe, la région mobile $\sigma 1.1$ se situe dans le canal de fixation de l'ADN (Mekler *et al.*, 2002) et la région 3.2 se situe à proximité du site actif (Murakami *et al.*, 2002a; Murakami *et al.*, 2002b; Vassilyev *et al.*, 2002). Dans ce complexe, les brins d'ADN ne sont pas encore séparés (Fig. 8 A et 9 D).

En utilisant des promoteurs différents (par exemple T7 A1, *lacUV5* et λ Pr), des complexes supposés être les premiers complexes formés lors de l'interaction ARN pol/ADN ont été mis en évidence en réalisant les incubations à basse température (entre 0°C et 10°C) (Fig. 8 A). Dans le cas de promoteurs comme T7 A1 ou *lacUV5*, l'ARN pol protège l'ADN des positions -55 à -6 (Kovacic, 1987; Schickor *et al.*, 1990; Mecsas *et al.*, 1991) et le complexe ainsi formé a été appelé RPc1 (Record *et al.*, 1996). De plus, une seule face de la double hélice est protégée par l'ARN pol (Schickor *et al.*, 1990).

A des températures plus élevées (entre 10 et 15°C), le complexe RPc1 va évoluer vers le complexe RPc2 dans lequel la double hélice n'est pas encore ouverte. Les protections de l'ADN générées par l'ARN pol sont différentes dans RPc2. En effet, la molécule d'ADN est protégée de l'action de la DNase I des régions allant de la position -55 à la position +20 par l'ARN pol (Kovacic, 1987; Schickor *et al.*, 1990; Mecsas *et al.*, 1991). Un changement de

conformation de l'ARN pol doit donc avoir lieu lors du passage R_{Pc1} vers R_{Pc2} (Fig. 8 B). L'ADN pénétrerait donc plus profondément à l'intérieur de son canal de fixation dans l'holoenzyme (Studitsky *et al.*, 2001), ce qui implique un déplacement de la région σ 1.1 (Fig. 8 B). Le changement de conformation de l'ARN pol serait majoritairement dû au mouvement de cette région 1.1 (Fig. 8 B) (Murakami et Darst, 2003). Le changement de conformation pourrait aussi être lié à un mouvement convergent des sous-unités β et β' (fermeture des « pinces ») (Craig *et al.*, 1998; Saecker *et al.*, 2002).

Dans le cas du promoteur λ Pr, le complexe intermédiaire formé à 0°C a été appelé I1 (Craig *et al.*, 1998). Dans ce complexe, l'ARN pol couvre une région plus large de l'ADN (de la position -40 à la position +30) par rapport au complexe R_{Pc1}. Ce complexe ressemblerait donc au complexe R_{Pc2} où les protections de l'ADN par l'ARN pol se situent des positions -55 à +20. La durée de vie de cet intermédiaire I1 est très courte même à faible température (Craig *et al.*, 1998). Ce complexe serait en équilibre avec le complexe I2. La formation de ce complexe I2 serait liée à un changement de conformation de l'ARN pol (pour revue : deHaseth *et al.*, 1998). Ce changement de structure pourrait correspondre comme dans le cas du passage de R_{Pc1} à R_{Pc2}, à un mouvement convergent des sous-unités β et β' (Craig *et al.*, 1998; Saecker *et al.*, 2002).

Au niveau des complexes que nous avons appelés intermédiaires (qui pourraient correspondre à R_{Pc2} ou I2) (Fig. 8 B), l'interaction entre les deux brins d'ADN pourrait être déstabilisée grâce aux acides aminés aromatiques Tyr 425, Tyr 430 et Trp 433 du facteur σ (région 2.3). Ces acides aminés, situés à proximité de l'ADN, peuvent interagir avec celui-ci lors de son d'ouverture transitoire (respiration), ce qui permettrait l'ouverture de la double hélice et entraînerait la formation du complexe R_{Po} (nucléation) (Murakami *et al.*, 2002b; Murakami et Darst, 2003).

Cette ouverture de la double hélice pourrait être un mécanisme séquentiel (Roy *et al.*, 2004): la première étape pourrait être un basculement de la base A-11 du brin non matrice, qui viendrait se loger entre deux acides aminés aromatiques Tyr 430 et Trp 433 (Helmann et deHaseth, 1999; Fenton *et al.*, 2000; Tsujikawa *et al.*, 2002). L'importance de cette base pour l'initiation de l'ouverture de la double hélice a été démontrée par son remplacement par une 2-aminopurine qui possède des propriétés de fluorescence (Tsujikawa *et al.*, 2002). En accord avec un basculement du résidu en position -11 au sein d'une poche hydrophobe de l'holoenzyme constituée par le domaine σ 2.3, il a été observé que 60% des 2-aminopurines sont immobilisées au sein de l'ARN pol (Tsujikawa *et al.*, 2002). Dans une deuxième étape,

l'ouverture de l'ADN se produirait au niveau des positions +2 et +3, grâce à la liaison des sous-unités β et β' (Saecker *et al.*, 2002). Ensuite dans une troisième étape, il y aurait séparation des brins des positions +1 à -10 (Roy *et al.*, 2004). Sur la figure 8 B est représenté le déroulement de l'ADN créé par l'ouverture de la double hélice. Ce déroulement pourrait par la suite rendre l'ADN de la bulle de transcription plus flexible permettant ainsi la courbure de l'ADN à l'entrée du site actif et ainsi rapprocher le brin matrice du site actif (Saecker *et al.*, 2002; Murakami et Darst, 2003).

Récemment, une technique d'empreinte par des radicaux hydroxyles générés par rayons X a été utilisée pour mettre en évidence les intermédiaires réactionnels au cours du processus de formation des complexes RPo à 37°C sur le promoteur fort T7 A1 (Sclavi *et al.*, 2004, communication personnelle). Les rayons X de forte intensité permettent la génération de radicaux hydroxyles à partir d'eau en un temps très court (quelques millisecondes) (Sclavi *et al.*, 1997). Cette technique peut être couplée à un système de mélange rapide. L'analyse des protections par l'ARN pol contre l'action des radicaux hydroxyles permet d'estimer le positionnement de l'ARN pol au niveau du promoteur et les variations de protections des différentes régions du promoteur en fonction du temps donnent des indications sur la cinétique de formation des complexes ARN pol/ADN.

Les résultats montrent qu'à 37°C, les contacts initiaux concernent la boîte -35 et les régions en amont de cette boîte respectivement par le domaine $\sigma 4.2$ et l' α CTD. Après le changement de structure de l'ARN pol, la zone de protection s'étend pour inclure la boîte -10. La nucléation de l'ouverture de la double hélice se produirait au niveau de ce complexe. La dernière étape serait la fixation de l'ADN entre les « pinces » β et β' (Sclavi *et al.*, 2004, communication personnelle). Le complexe ainsi formé qui serait équivalent au complexe RPo pourrait permettre l'initiation de la transcription. Les étapes qui permettent la formation du complexe RPo à 37°C seraient différentes des étapes décrites auparavant (RPc1, RPc2). Cette différence pourrait être due au fait qu'à 37°C, l'ouverture de la double hélice est facilitée. La nucléation de l'ouverture se produirait très rapidement après le contact de l'ARN pol au niveau de la boîte -10.

I.2.1.2) Le complexe RPo.

L'ouverture complète de la double hélice est obtenue dans le complexe RPo où le site +1 d'initiation de la transcription se trouve sous forme simple brin (Fig. 8 C et 9 C et D). Le passage des complexes RPc au complexe RPo est souvent une étape limitante de l'initiation de la transcription.

En effet, pour un grand nombre de promoteurs, la formation du complexe RPo est l'étape limitante. Parfois, le complexe formé est instable comme c'est le cas pour le promoteur *rrnB* P1 des opérons ribosomiques d'*E. coli* mais dans la majorité des cas, il est stable. Une propriété des complexes RPo est leur résistance à l'action de l'héparine, un polyanion qui est un compétiteur non spécifique de l'ADN. Cette résistance est utilisée pour mesurer le taux de complexes RPo formés et pour étudier ces complexes. Par traitement à l'héparine, les complexes aspécifiques ADN-ARN pol et les complexes RPc sont dissociés.

Le motif TG présent dans les promoteurs étendus jouerait aussi un rôle dans la formation des complexes RPo, ceci a été démontré au niveau du promoteur *galP1* (Burns *et al.*, 1999). Il jouerait aussi un rôle dans la stabilité de ces complexes, comme cela a été montré pour le promoteur *amyP* de *Bacillus subtilis* (Voskuil et Chambliss, 2002).

Au niveau structural, les acides aminés aromatiques Tyr 430, Tyr425, Trp 433 et Trp 434 de la région $\sigma 2.3$ interagiraient avec le simple brin non matrice de la molécule d'ADN (Marr et Roberts, 1997; Fenton *et al.*, 2000; Murakami *et al.*, 2002b) (Fig. 11), ce qui stabiliserait la bulle de transcription, dont la formation est initiée par le basculement de la base A-11. Les résidus en simple brin de la boîte -10 qui jouent un rôle critique dans l'interaction avec l'holoenzyme au sein du complexe RPo sont situés aux positions -11, -13, -14 et -15 (tAtAAT). La base à la position -11 est donc à la fois importante pour la fixation de l'ADN double brin dans les complexes fermés mais aussi dans l'interaction de la région d'ADN simple brin avec l'ARN pol dans le complexe RPo (Helmann et deHaseth, 1999; Fenton et Gralla, 2001). En étudiant les propriétés de fixation d'ADN en fourche (Fig. 9A), l'équipe de Gralla a montré que l'ARN pol se lie de manière très stable avec un ADN en double brin jusqu'à la position -12, le résidu -11 étant en simple brin (Guo et Gralla, 1998). Les complexes formés sont résistants à l'héparine indiquant que le complexe formé possède certaines caractéristiques du complexe RPo (Guo et Gralla, 1998). Les auteurs concluent que la paire de base à la position -12 constitue l'extrémité de la bulle de transcription et n'est pas ouverte au cours de la formation du complexe RPo lors de l'initiation. Ces auteurs ont confirmé l'importance de la base -11 en simple brin (Guo et Gralla, 1998). Le rôle de la paire

de base à la position -12 est en accord avec le modèle 3D proposé pour l'ARN pol associée avec un ADN en fourche (Murakami *et al.*, 2002b). La paire de base -12 qui est fortement impliquée dans la liaison de l'ARN pol à l'ADN double brin ne serait pas ouverte dans le complexe RPo et l'acide aminé Trp 433 serait situé en face de cette paire de bases (Murakami *et al.*, 2002b).

La structure de l'holoenzyme associée au fragment d'ADN en fourche montre en particulier qu'un domaine de β' (BZD) pourrait entrer en contact avec l'ADN de la région interboîte à la position -22 sur le brin matrice et à la position -27 sur le brin non matrice (Fig. 9 B). Dans cette région interboîte, l'ADN qui se trouve au contact de l'ARN pol est courbé (courbure de 8°). Une courbure de plus forte amplitude (37°) est observée au niveau du motif TG (Fig. 9). Cette structure montre aussi que les acides aminés du facteur σ sont positionnés de façon à stabiliser le complexe ouvert.

Le modèle de complexe RPo construit par Murakami *et al.* (2002b) qui contient toute la séquence ADN allant des positions -60 à +25 (Fig. 9 C et D) montre que les deux brins d'ADN suivent deux tunnels différents (cf. Fig. 7) : pour le brin matrice, ce tunnel est formé par les sous-unités β , β' et par σ^{70} . Il passe à proximité du site actif (matérialisé par la position de l'ion Mg^{2+}) (Fig. 9) (Young *et al.*, 2002). Le brin non matrice est contacté jusqu'à la position -7 par le facteur σ , position au niveau de laquelle il change d'orientation (Fig. 9 C). Ainsi, les bases des positions -2 à +4 du brin non matrice passeraient dans un autre tunnel de l'ARN pol formé par deux lobes de la sous-unité β (β_1 et β_2) (Young *et al.*, 2002; Murakami *et al.*, 2002b). A la sortie de ces deux tunnels, les brins sont réappariés pour reformer une double hélice d'ADN qui est contactée par les sous-unités β et β' de l'ARN pol.

Au niveau du complexe RPo, lors de la fixation de l'ARN pol sur l'ADN de promoteurs, la DNase I clive plus efficacement l'ADN aux positions -45, -33 (-34) et -25 par rapport au site +1 (Ozoline et Tsyganov, 1995) (Fig. 9 D). Ces sites sont centrés au niveau du petit sillon de l'hélice d'ADN, qui se trouvent exposés lors de la fixation de l'ADN sur l'ARN pol (Murakami *et al.*, 2002b). L'exposition de ces sillons à la DNase I serait liée aux courbures induites par l'ARN pol : les courbures de 8° de la région interboîte et 36° au niveau de la boîte -35 vers l'ARN pol pourraient favoriser les clivages aux positions -25 et -33 alors que la liaison de l' α -CTD sur la séquence en amont de la boîte -35 pourrait favoriser le clivage à la position -45 (Murakami *et al.*, 2002b). L'apparition de ces clivages à la DNase I renforce l'idée de l'enroulement de l'ADN autour de l'holoenzyme. Cet enroulement autour de l'ARN

pol pourrait favoriser la séparation des deux brins d'ADN lors de la formation des complexes RPo (Coulombe et Burton, 1999).

I.2.1.3) Processus de formation du complexe RPelong.

L'enchaînement des premiers ribonucléotides commence dans le complexe RPo, ce qui conduit au complexe RPinit (Fig. 8 D et E). Des fragments d'ARN de petite taille (de 2 à 15 nucléotides suivant les promoteurs) sont synthétisés par l'ARN pol, qui peuvent, soit être rallongés jusqu'à donner des transcrits de grande taille permettant la synthèse des protéines, soit être relargués par l'enzyme. Ces petits transcrits relargués ont été appelés transcrits abortifs (Carpousis et Gralla, 1980) et ce mécanisme a été appelé transcription abortive (pour revue : Hsu, 2002b). Ce relargage de transcrits abortifs est un phénomène universellement conservé pour toutes les polymérases étudiées *in vitro*. Ces transcrits abortifs peuvent être majoritaires par rapport aux transcrits de grande taille (Hsu *et al.*, 2003). Chaque ARN pol passe par des étapes de transcription abortive avant d'entrer en élongation (Vo *et al.*, 2003a). Deux types de complexes RPo pourraient exister au cours de l'initiation de la transcription (Kubori et Shimamoto, 1996; Vo *et al.*, 2003a). *In vitro*, certains complexes appelés non productifs (ou moribonds) permettraient uniquement la synthèse de transcrits abortifs, alors que les complexes appelés productifs permettent la synthèse de grands transcrits (Kubori et Shimamoto, 1996; Hsu, 2002b; Vo *et al.*, 2003a). Les complexes productifs produisent aussi des transcrits abortifs au cours du mécanisme, mais au bout de quelques cycles, permettent tout de même la synthèse d'un grand transcrit (Vo *et al.*, 2003a).

Lorsque le transcrit abortif est libéré de l'ARN pol, il y a retour du complexe ARN pol/ADN au stade du complexe RPo. Ce complexe peut réinitier l'enchaînement de nucléotides (ce cycle peut se produire plusieurs fois) jusqu'à ce qu'un complexe productif entre en phase d'élongation (complexe RPelong) (Fig. 8 F).

Plusieurs modèles tentent d'expliquer ce phénomène de transcription abortive. Le premier fait appel à la structure de l'holoenzyme (Murakami *et al.*, 2002a; Vassylyev *et al.*, 2002) qui montre que la molécule d'ARN en cours d'élongation pourrait entrer en interaction avec le domaine $\sigma 3.2$ qui bloque le tunnel de sortie de l'ARN (Fig. 8 C et D). Dans ce cas, la taille maximale de la molécule d'ARN avant d'être bloquée serait de 12 nucléotides, ce qui correspond approximativement à la taille des transcrits abortifs. Deux possibilités pourraient se présenter à la molécule d'ARN venant de débiter sa synthèse. Le domaine $\sigma 3.2$ peut être déplacé, ce qui permet l'augmentation de la taille de l'ARN et donc la synthèse d'un transcrit

de grande taille. Dans le cas contraire, la région $\sigma_{3.2}$ ne peut pas être déplacée et le transcrit ne peut pas être rallongé. Ce dernier se décroche donc de la polymérase, qui revient sous la forme d'un complexe RPo. Une nouvelle synthèse commence alors, jusqu'à ce que la sous-unité $\sigma_{3.2}$ soit déplacée. Une fois que la région 3.2 est déplacée, l'interaction existant entre σ_4 et l'ARN polymérase pourrait être défavorisée, ce qui entraînerait une baisse de l'affinité pour le facteur σ qui se décrocherait de l'ARN pol. La polymérase se décrochant glisserait alors le long de l'ADN du promoteur et l'étape d'élongation (RPelong) débiterait (Fig. 8 E et F) (Murakami et Darst, 2003). Selon DeHaseth *et al.*, (1998), le déplacement du domaine $\sigma_{3.2}$ pourrait se faire du fait de la stabilité de l'hybride ADN-ARN. Un des faits n'étant pas expliqué par ce modèle est que la taille des transcrits abortifs peut varier de 2 à 15 nucléotides, alors que le relargage des transcrits abortifs par cet encombrement stérique ne se produirait qu'à une taille constante du transcrit.

Un deuxième modèle a été décrit par Hsu en 2002 qui permet lui d'expliquer les différences de taille des transcrits abortifs (Hsu, 2002b). Ce modèle fait intervenir des intermédiaires où la molécule d'ADN se trouve sous une forme « stressée ». Ces intermédiaires sont formés à partir du complexe RPo par ajout successif de nucléotides. En effet, lorsque les deux nucléotides initiateurs (1 et 2) sont ajoutés au complexe RPo au niveau des sites i et $i+1$, le site actif de l'enzyme doit se déplacer, laissant le site $i+1$ accessible afin d'y inclure le troisième nucléotide. Les contacts au niveau de la boîte -10 et de la boîte -35 sont alors toujours présents. Ce déplacement entraîne l'ouverture de la première paire de base à l'extrémité de la bulle de transcription et impose des contraintes à l'ADN du fait de son repliement à l'intérieur de l'enzyme. Pour relâcher une partie de ces contraintes, la distorsion se propage aux brins d'ADN en amont ou en aval du site actif. Si le déplacement se fait en aval, l'ARN synthétisé peut se décrocher du brin matrice et est relargué par la polymérase. Dans le cas inverse, si le déplacement se fait en amont, ce mouvement pourrait favoriser le déplacement de la polymérase vers la position suivante (quatrième nucléotide). Une fois que ce quatrième nucléotide est ajouté, le site actif se déplace de nouveau, augmentant les contraintes au niveau de l'ADN. A ce niveau, la distorsion peut à nouveau se propager soit en amont, favorisant ainsi le déplacement vers la position 5, soit en aval, favorisant le relargage du transcrit. A chaque position, le transcrit peut donc être soit allongé soit relargué. Lorsque la longueur de l'hétéroduplex est de 8 à 9 nucléotides, l'énergie apportée par les contraintes imposées à l'ADN est suffisante pour réduire la bulle de transcription en reformant les paires de bases au niveau de la boîte -10. Les interactions entre σ_2 et la boîte -10 sont donc

modifiées, ce qui pourrait entraîner le décrochage du facteur sigma et donc la formation du complexe RPelong.

Ce modèle permet donc d'expliquer les tailles différentes des transcrits abortifs mais aussi d'expliquer pourquoi les positions de formation des complexes RPelong (fin de la transcription abortive) changent en fonction des promoteurs (Hsu *et al.*, 2003).

Cette étape d'échappement du promoteur est une étape limitante pour certains promoteurs comme *lacUV5*, T5 N25 (Kammerer *et al.*, 1986) et *malT* (Eichenberger *et al.*, 1997).

1.2.1.4) Le complexe RPelong et mécanisme d'élongation.

Des analyses par empreinte à la DNase I montrent que les protections de l'ADN par l'ARN polymérase au niveau du complexe RPelong se situent entre les positions -5 et +25 (Straney et Crothers, 1987) indiquant un déplacement de l'holoenzyme le long de l'ADN. A partir de la structure de l'apoenzyme de *T. aquaticus* (Zhang *et al.*, 1999) et de résultats de pontage entre l'ADN, l'ARN en cours d'élongation et les sous-unités de l'ARN pol (Korzheva *et al.*, 2000), un modèle du complexe RPelong a été établi. Ce modèle est schématisé dans la figure 8 F. Cette représentation est aussi basée sur les résultats d'études cristallographiques réalisées sur des ARN pol isolées d'autres organismes (eucaryotes et archaea) (Bartlett *et al.*, 2000; Gnatt *et al.*, 2001).

Au niveau du complexe RPelong, l'ARN pol devient très processive du fait de la forte stabilité de ces complexes, ce qui permet la synthèse de transcrits de longue taille.

Comme indiqué précédemment, le départ à ce stade du facteur σ serait lié à la fin de la transcription abortive mais aussi au changement de structure de l'holoenzyme (Fig. 8 E et F). En 2001, ce fait a été remis en cause car le facteur sigma a été retrouvé en partie associé à l'ARN polymérase en cours d'élongation (Bar-Nahum et Nudler, 2001; Mukhopadhyay *et al.*, 2001). Le facteur σ n'est retrouvé associé tout au long du processus de transcription que dans 30% des cas, et la population d'ARN pol ayant retenu le facteur σ est distincte de celle ne l'ayant pas retenu. Cette population d'ARN pol pourrait réinitier la transcription plus facilement au niveau de certains promoteurs car l'étape de fixation de σ sur l'ARN pol pourrait constituer une étape limitante (Bar-Nahum et Nudler, 2001). De plus, le facteur σ qui reste associé à l'ARN pol durant l'élongation serait capable de reconnaître des boîtes -10 étendues qui pourraient se trouver dans la séquence ADN codant la protéine, ce qui entraînerait une pause de l'ARN pol à cet endroit (Ring *et al.*, 1996; Mooney et Landick,

2003). Un exemple a été décrit pour le promoteur *lacUV5* (Brodolin *et al.*, 2004). Les auteurs ont montré que le facteur σ induisait une pause de l'ARN pol au cours de l'élongation (position +17 par rapport au site +1 d'initiation de la transcription) indiquant que ce facteur restait associé à l'apoenzyme. Les auteurs ont démontré que cette pause était bien due au facteur σ qui reconnaissait une séquence proche du consensus pour une boîte -10. Par contre, les interactions entre l'apoenzyme et le facteur σ seraient plus faibles que dans l'holoenzyme, ce qui permettrait le changement de structure de l'ARN pol pendant l'élongation (Brodolin *et al.*, 2004).

Au cours de la phase d'élongation, des études réalisées sur l'ARN pol issue du phage T7 montrent un mécanisme d'addition des nucléotides qui semble être valable pour toutes les polymérases notamment les ARN pol bactériennes (Yin et Steitz, 2004; Temiakov *et al.*, 2004; Landick, 2004) (Fig. 13). Ce mécanisme est basé sur la présence de deux ions Mg^{2+} catalytiques. La présence de ces deux ions a été confirmée pour les ARN pol bactériennes. Le site actif de l'ARN pol existe sous deux formes : une forme ouverte et une forme fermée. La différence est due au mouvement d'une hélice α de l'ARN pol (hélice O) et d'un acide aminé Tyr 639 (Y639) de cette hélice O, qui stabilise la dernière paire de bases de l'hybride ADN-ARN (au niveau du site P). Le nucléotide pénètre ensuite dans le site actif au niveau du site S. L'hélice O se déplace ensuite ainsi que l'acide aminé Tyr 639. Ce déplacement permet à la catalyse d'avoir lieu. Une fois que le PPI est relargué dans le milieu, l'ARN pol se déplace d'un nucléotide (translocation) et l'hélice revient dans sa forme ouverte. La figure 13 B montre que le site actif de l'ARN pol bactérienne serait organisé de la même manière que le site actif de l'ARN pol T7, confirmant que le mécanisme serait identique pour ces deux enzymes.

I.2.1.5) La terminaison de la transcription.

La transcription s'arrête lorsque la polymérase rencontre un site d'arrêt de la transcription qui peut être de deux types : séquence-dépendant ou facteur-dépendant (pour revue Richardson et Greenblatt, 1996; Nudler et Gottesman, 2002).

Dans le cas de la terminaison de la transcription séquence-dépendante, le signal d'arrêt est constitué par la formation sur le transcrit d'une structure tige boucle riche en paires de base G-C suivie de 7 à 9 résidus U. La séquence de U induit une brève pause de l'ARN pol en

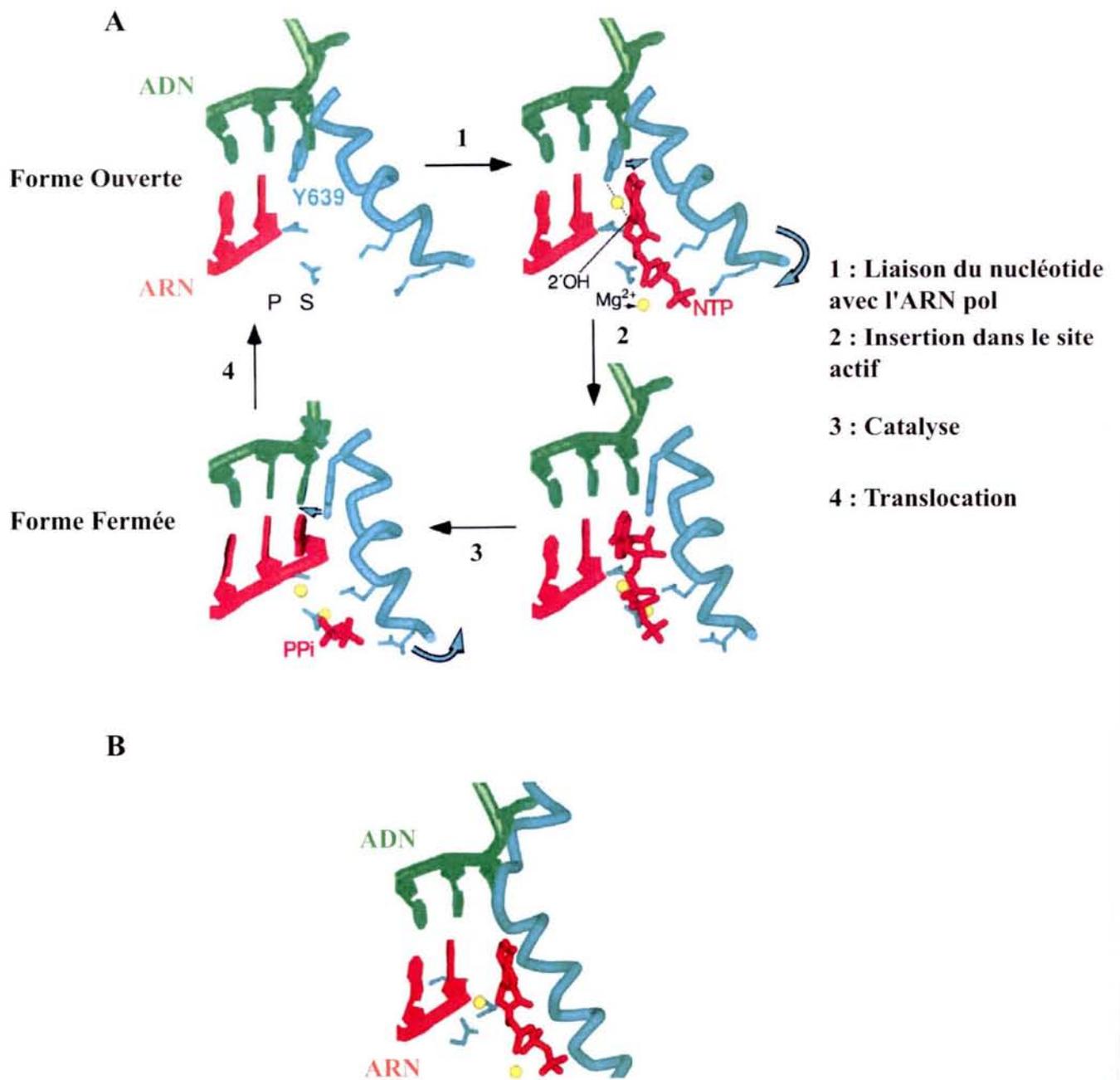


Figure 13 : Structure du site actif des ARN pol (d'après Landick, 2004 ; Yin *et al.*, 2004 et Temiakov *et al.*, 2004).

A) Représentation du site actif de l'ARN pol T7 et des quatre étapes nécessaires à l'addition d'un nucléotide (Landick, 2004). La molécule d'ADN en simple brin est représentée en vert, la molécule d'ARN nouvellement synthétisée en rouge et l'hélice-O de l'ARN pol en bleu. Cette hélice serait présente dans toutes les ARN pol et son mouvement serait essentiel à la catalyse. Les deux points jaunes représentent les deux ions catalytiques Mg^{++} impliqués dans le mécanisme. Les sites P et S représentent le site produit (site i) et le site substrat (site $i+1$).

B) Représentation du site actif de l'ARN pol bactérienne

cours d'élongation, ce qui laisse le temps à la structure tige boucle de se former. Le passage de l'état de pause à l'état de terminaison est lié à la stabilité de la tige boucle formée qui déstabilise certaines des interactions ARN pol-ARN. Par ailleurs, l'hybride dA/U formé au niveau de la série de U enchaînés après synthèse de la tige/boucle sont peu stables, ce qui favorise le décrochement de l'ARN naissant du brin d'ADN matrice (Nudler et Gottesman, 2002).

Dans le cas de la terminaison facteur-dépendante, l'arrêt de la transcription est liée à la fixation de facteurs protéiques sur l'ARN synthétisé. Deux de ces facteurs ont été identifiés chez *E. coli* : il s'agit du facteur Rho qui agit au niveau d'un grand nombre de sites sur le chromosome bactérien et du facteur MFD (Mutation Frequency Decline) qui permet de décrocher les ARN pol bloquées au niveau de lésions de l'ADN induites par les UV. La protéine Rho (sous forme d'anneau hexamérique) se fixe sur des séquences au niveau de l'ARN appelées RUT (Rho-utilization site). Rho se déplace ensuite le long de l'ARN en hydrolysant des molécules d'ATP et interagit avec l'ARN pol qui fait une pause au cours de l'élongation. Cette interaction permettrait de décrocher l'ARN synthétisé (pour revue : Nudler et Gottesman, 2002).

I.2.2) Mécanisme pour l'holoenzyme $E\sigma^{54}$.

La figure 14 résume le mécanisme d'initiation de la transcription pour cette holoenzyme. La différence par rapport au mécanisme de $E\sigma^{70}$ est essentiellement la nécessité de la fixation d'un activateur en amont du promoteur et de l'hydrolyse d'une molécule d'ATP par l'activité ATPase de l'activateur, afin d'aboutir à la formation d'un complexe ouvert (pour revue : Buck *et al.*, 2000).

La première étape est la fixation de l'holoenzyme sur la séquence promoteur située en amont du gène à transcrire. Un complexe fermé est ainsi obtenu. La particularité de ce complexe est sa grande stabilité. Il ne peut donc pas évoluer vers un complexe ouvert comme pour l'holoenzyme $E\sigma^{70}$. La fixation d'un activateur en amont du promoteur, ainsi que l'hydrolyse d'une molécule d'ATP par le domaine ATPase de l'activateur permet à l'holoenzyme d'ouvrir la double hélice d'ADN et ainsi d'obtenir un complexe ouvert RPo. La distance de fixation peut être très grande et la particularité de ce type d'activateur est la formation d'une boucle d'ADN permettant la liaison de l'activateur et de l'ARN pol (Su *et al.*, 1990). L'hydrolyse de

la molécule d'ATP est l'élément déclencheur qui permet le passage RPc à RPo (Wedel et Kustu, 1995) mais le mécanisme d'activation n'est pas connu actuellement. En fonction des signaux extérieurs, l'activité ATPase de l'activateur est régulée, ce qui influence l'initiation de la transcription. Une fois le complexe RPo obtenu, les premiers nucléotides sont enchainés et la transcription peut alors commencer.

Un exemple est l'initiation de la transcription par $E\sigma^{54}$ sous l'influence de l'activateur de la transcription NtrC qui se trouve sous forme dimérique en solution (Fig. 14). Pour cet activateur de la transcription et plus généralement pour un grand nombre d'activateur de la transcription de l'holoenzyme $E\sigma^{54}$, l'activité ATPase dépend du taux de phosphorylation de NtrC, et ce taux est régulé en fonction des conditions physiologiques du milieu extérieur (Rombel *et al.*, 1998). Un acide aspartique de l'extrémité N-terminale de NtrC est phosphorylé par la protéine NtrB dans des conditions de limitation en azote et cette phosphorylation entraîne un changement de structure de l'activateur, ce qui augmente son affinité pour l'ADN. NtrC-P se fixe sur son opérateur qui est composé de deux sites et multimérise, alors que la forme dimérique non phosphorylée ne peut se fixer sur un seul site de liaison et ne multimérise pas (Rippe *et al.*, 1998) (Fig. 14). Le multimère formé sur l'ADN acquiert l'activité ATPase (Weiss *et al.*, 1991) qui se situe au centre de ce complexe. Après l'hydrolyse d'une molécule d'ATP, la double hélice d'ADN est ouverte, puis la polymérase entre en élongation. Le système NtrB et NtrC fait partie des systèmes à deux composants qui seront décrits dans la deuxième partie de l'introduction.

II.) Régulation de l'initiation de la transcription.

Différents mécanismes permettent le contrôle de la transcription chez *E. coli*.

Le premier niveau de contrôle est le choix du facteur σ associé à l'apoenzyme puisque ce facteur définit les séquences promotrices reconnues. La production des différents facteurs σ est influencée par les conditions environnementales ou physiologiques. La séquence des promoteurs est aussi un paramètre important pour le contrôle de l'initiation de la transcription. Les variations des conditions environnementales peuvent aussi changer la conformation (courbure ou surenroulement) et l'accessibilité des promoteurs au sein de l'ADN, ceci sous l'action de facteurs protéiques se liant à l'ADN.

Le deuxième niveau de contrôle implique la régulation par des facteurs agissant en *trans* qui reconnaissent des séquences sur l'ADN et d'éléments agissant en *cis*, ce qui permet de modifier les paramètres thermodynamiques et cinétiques des différentes étapes de l'initiation de la transcription, en particulier les constantes k_1 et k_2 relatives à la fixation de l'holoenzyme sur le promoteur, les constantes k_3 et k_{-3} relatives à la formation et la stabilité du complexe RPo mais aussi l'échappement de l'ARN pol de sa région promotrice. L'influence de ces facteurs peut être positive ou négative suivant qu'ils sont activateurs ou repréresseurs.

II.1) Contrôle de l'initiation de la transcription par le choix du facteur σ .

Une cellule d'*E. coli* contient approximativement 2000 molécules d'ARN pol assemblées sous forme d'apoenzymes (Ishihama, 1981) pour environ 4000 gènes (Blattner *et al.*, 1997). L'identité des facteurs σ présents dans ces apoenzymes est donc très importante pour permettre à la cellule de ne transcrire que les gènes adaptés dont les produits sont requis dans les conditions environnementales où elle se trouve. Un changement de facteur σ permet de modifier rapidement le profil de transcription et donc une adaptation rapide aux conditions extérieures. Le choix du facteur σ dépend de son affinité relative pour l'apoenzyme mais aussi de la présence de facteurs antisigma qui empêchent les facteurs σ de se fixer sur l'apoenzyme (pour revue : Hughes et Mathee, 1998; Wosten, 1998; Hengge-Aronis, 2002). Il existe donc une compétition entre les différents facteurs σ pour la formation des différentes holoenzymes. En effet, le facteur σ^{70} est le plus exprimé quelque soit la période de croissance et les conditions environnementales (Ishihama, 2000). Le facteur σ^{54} est synthétisé à un taux constant et faible dans la cellule (Jishage *et al.*, 1996). Par contre, les autres facteurs σ ne sont exprimés que dans des conditions particulières de croissance.

II.1.1) Le facteur σ^S .

Le facteur σ^S est considéré comme un facteur général de réponse à différents stress (pour revue, Hengge-Aronis, 2002). La quantité du facteur σ^S est fortement augmentée dans des conditions de stress acide (Lee *et al.*, 1995), thermique (Sledjeski *et al.*, 1996; Muffler *et al.*, 1997a) ou hyperosmotique (Lange et Hengge-Aronis, 1994; Hengge-Aronis, 1996) mais aussi lorsque la bactérie entre en phase stationnaire de croissance, lorsque la concentration en

nutriments est limitante et que le taux de division cellulaire est faible (Lange et Hengge-Aronis, 1991) (Fig. 15). La synthèse du facteur σ^S assure à la bactérie une capacité de résistance à des conditions de stress, en particulier, il est essentiel pour la survie d'*E. coli* lors de la phase stationnaire de croissance.

L'augmentation de la quantité du facteur σ^S en phase stationnaire de croissance ou en condition de stress est régulée à trois niveaux et ceci par un grand nombre de facteurs (Fig. 15). Le premier niveau de régulation est la transcription du gène *rpoS* codant le facteur σ^S (Lange et Hengge-Aronis, 1991). Lors de l'entrée en phase stationnaire de croissance, la transcription du gène *rpoS* est activée d'un facteur cinq à dix (Lange et Hengge-Aronis, 1991; Lange et Hengge-Aronis, 1994). Le mécanisme déclencheur serait la diminution de la vitesse de croissance (Fig. 15) (Lange et Hengge-Aronis, 1991).

Le deuxième niveau de régulation se situe à l'étape de traduction de l'ARNm *rpoS* qui est influencée par plusieurs facteurs notamment les petits ARN régulateurs DsrA (Sledjeski *et al.*, 1996) et OxyS (Zhang *et al.*, 1998a), des protéines de liaison à l'ADN comme H-NS (Yamashino *et al.*, 1995) et HU($\alpha\beta$) (Balandina *et al.*, 2001), la protéine de liaison à l'ARN Hfq (Muffler *et al.*, 1996) et des métabolites comme l'UDP-Glucose (pour revue : Hengge-Aronis, 2002). La structure de la région 5' non traduite de l'ARNm existerait sous 2 formes : l'une compatible avec la traduction et l'autre non. Les facteurs cités précédemment pourraient influencer cette structuration différentielle (Hengge-Aronis, 2002). Les différentes conditions de stress indiquées dans la figure 15 influencent les effets des facteurs cités ci-avant, ce qui permet une traduction fortement régulée.

Le dernier niveau de régulation est la dégradation de ce facteur σ , qui en conditions normales de croissance est très rapide. La dégradation est assurée par la protéase ClpXP et par la protéine RssB qui dirige σ^S vers la protéolyse (pour revue : Hengge-Aronis, 2002). La protéine RssB sous sa forme phosphorylée interagit avec la sous-région 3.0 du facteur σ^S , ce qui libère l'extrémité N Terminale de ce facteur qui interagit avec la protéase ClpXP (Studemann *et al.*, 2003). En condition de stress, la dégradation de σ^S est fortement ralentie et différents facteurs ont une influence sur le taux de phosphorylation de RssB, sur sa quantité dans la cellule ou sur l'activité de la protéase ClpXP (Hengge-Aronis, 2002).

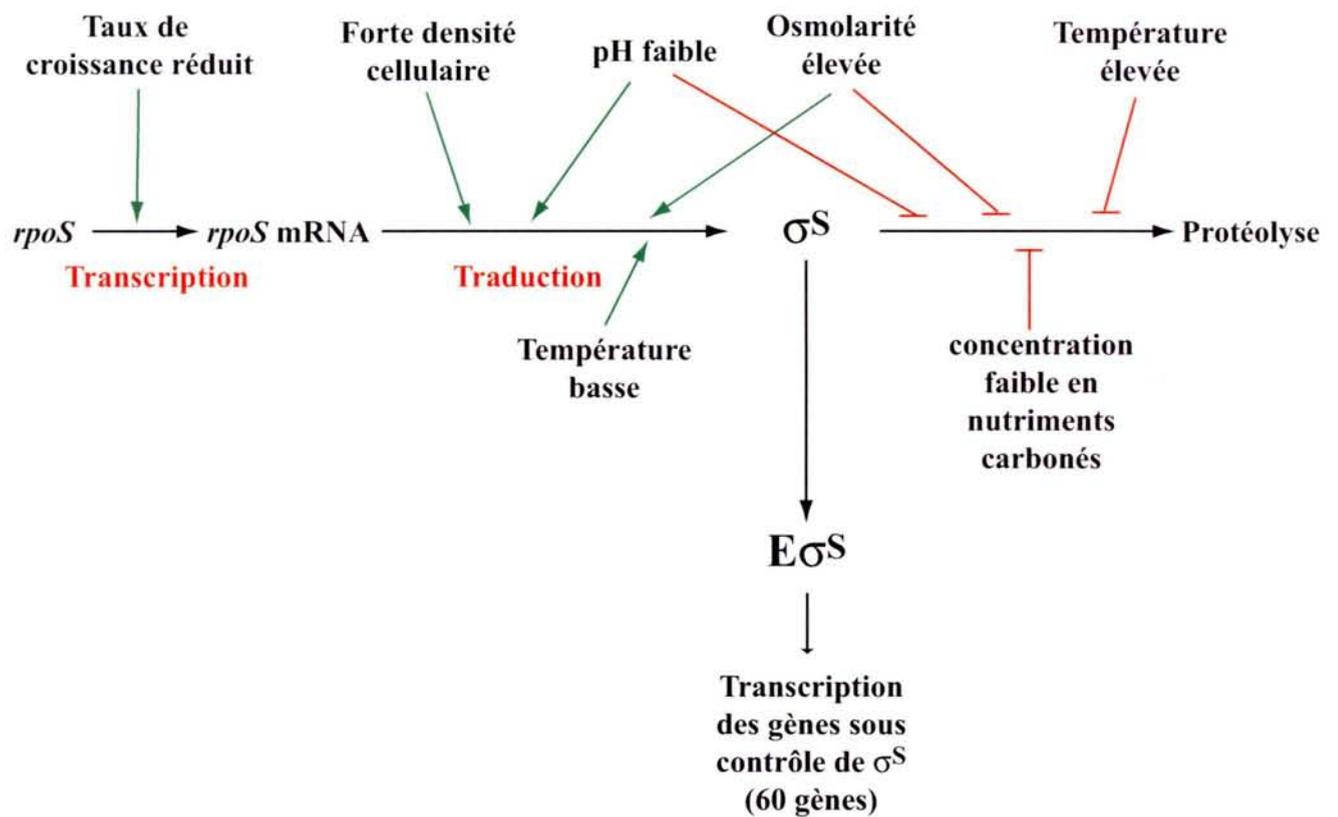


Figure 15 : Les différents types de régulation observés pour la transcription, la traduction et la dégradation du facteur σ^S (38) (d'après Hengge-Aronis, 2002).

Les étapes de la synthèse du facteur σ^{70} sont régulées par un grand nombre de facteurs extérieurs.

Les flèches vertes indiquent l'activation de l'étape alors que les traits rouges indiquent les inhibitions.

$E\sigma^S$ représente l'holoenzyme associée au facteur σ^S .

II.1.2) Le facteur σ^{32} .

Le facteur σ^{32} est activé lors de la présence de protéines dénaturées dans le cytoplasme qui sont notamment abondantes lors d'un choc thermique. Il permet d'initier la transcription des gènes codant des protéines de choc thermique (HSP : Heat Shock Protein) (pour revue : Mager et De Kruijff, 1995). Les protéines chaperonnes DnaK, DnaJ et GrpE sont des protéines HSP. Elles permettent la renaturation des protéines dans le cytoplasme.

La quantité du facteur σ^{32} dans la cellule dépend principalement comme pour σ^S de la régulation des étapes d'initiation de la transcription de son gène, de la traduction de son ARNm et de sa dégradation (pour revue : Mager et De Kruijff, 1995). Comme le facteur σ^S , le facteur σ^{32} est dégradé rapidement. La dégradation est liée à l'association des protéines chaperonnes DnaK, DnaJ et GrpE sur le facteur σ^{32} (Gamer *et al.*, 1996) puis à la dégradation par certaines protéases comme FtsH (Tomoyasu *et al.*, 1995). Lors de la dénaturation de protéines dans le cytoplasme, ces protéines ne se fixent plus sur σ^{32} mais sur les protéines dénaturées (Gamer *et al.*, 1996). La stabilité du facteur σ^{32} est alors augmentée, ce qui favorise son association avec l'apoenzyme et la transcription des gènes appartenant au régulon sous le contrôle de ce facteur. L'action de σ^{32} est limitée par le fait que σ^{32} est nécessaire à la transcription des protéines DnaK, DnaJ et GrpE qui inhibent son activité.

II.1.3) Le facteur σ^{28} .

Le facteur σ^{28} , encore appelé FliA, est impliqué dans la transcription des gènes de synthèse des flagelles (Arnosti et Chamberlin, 1989). La protéine FlgM est un facteur antisigma qui se fixe sur FliA et empêche la fixation de σ^{28} sur l'apoenzyme (Gillen et Hughes, 1991; Sorenson *et al.*, 2004) (pour revue : Hughes et Mathee, 1998). Lorsque la synthèse du flagelle est terminée, cette protéine est excrétée par le flagelle (Kutsukake, 1994; Hughes et Mathee, 1998). Le facteur σ^{28} ainsi libéré est disponible pour l'assemblage avec l'apoenzyme ($E\sigma^{28}$) et l'initiation de la transcription des gènes codant par exemple la protéine de constitution du flagelle FliC (Kundu *et al.*, 1997).

II.1.4) Les facteurs $\sigma^{24(E)}$ et σ^{FecI} .

Ces facteurs ne sont synthétisés que dans des conditions très particulières : en cas de choc thermique périplasmique pour $\sigma^{24(E)}$ et par la présence de citrate de fer dans le milieu pour $\sigma^{19(FecI)}$.

Les mécanismes d'activation de ces facteurs σ sont connus. Pour σ^E , des protéines dénaturées dans le périplasme, par exemple des porines, activent les protéases membranaires DegS et YaeL (Walsh *et al.*, 2003). Ces protéases clivent de manière séquentielle le facteur transmembranaire antisigma RseA, qui est associé au facteur σ^E et le séquestre au niveau de la membrane (Kanehara *et al.*, 2002; Campbell *et al.*, 2003). Le facteur σ^E est alors libéré dans le cytoplasme, il s'associe à l'apoenzyme pour activer la transcription des gènes sous son contrôle, notamment des protéines chaperonnes permettant de renaturer les porines (pour revue : Ades, 2004).

Pour σ^{FecI} , la présence de citrate de fer dans le milieu extérieur est détectée par le transporteur FecA localisé dans la membrane externe. Après la liaison du citrate de fer, la structure 3D de FecA est modifiée (Ferguson *et al.*, 2002). Ce signal est ensuite transmis à FecR qui est situé dans la membrane cytoplasmique, ceci grâce à l'interaction du segment N-terminal de FecA et du segment C-terminal de FecR (Kim *et al.*, 1997; Welz et Braun, 1998). L'autre extrémité de FecR se situe dans le cytoplasme et cette séquence protéique interagit avec le facteur σ^{FecI} . Le facteur σ est alors activé par un mécanisme encore inconnu, ce qui conduit à sa liaison à l'apoenzyme pour initier la transcription des gènes *fecABCDE*. Cet ensemble de gènes code des enzymes responsables de la synthèse des protéines permettant le transport du fer à l'intérieur de la cellule (Braun, 1997). Ce mécanisme est régulé par la protéine Fur, qui inhibe la transcription de l'opéron *fecABCDE* lorsque du fer est présent en excès dans la cellule (cf partie régulation en *trans*) (pour revue : Braun *et al.*, 2003).

II.1.5) Conséquences des variations de concentration des facteurs σ sur le taux des différentes holoenzymes dans la cellule.

L'affinité pour l'apoenzyme des différents facteurs σ chez *E. coli* a été mesurée *in vitro* (Tableau 3) (Maeda *et al.*, 2000a). Les auteurs ont surproduit chacun des sept facteurs σ et ont mesuré la constante de dissociation (Kd) du complexe ARN pol/ σ pour chaque facteur

(Maeda *et al.*, 2000a). De plus, les auteurs ont confirmé ces résultats en réalisant des compétitions de fixation des différents facteurs σ sur l'apoenzyme (Maeda *et al.*, 2000a).

σ	Kd (nM)	Nombre de molécules par cellule	Taux d'holoenzyme (%)
σ^{70}	0,26	700	78
σ^{54}	0,30	110	8
σ^S	4,26	<1	0
σ^{32}	1,24	<10	0
σ^{28}	0,74	350	14
σ^{24}	2,43	<10	0
σ^{FecI}	1,73	<1	0

Tableau 3 : Mesure de l'affinité *in vitro* des facteurs σ pour l'apoenzyme (colonne 1) et estimation du nombre de molécules des différents facteurs σ dans la cellule (colonne 2) (en phase exponentielle de croissance). Le taux d'holoenzyme a été calculé à partir de ces deux valeurs (colonne 3) (Jishage et Ishihama, 1995; Jishage *et al.*, 1996; Maeda *et al.*, 2000b; Ishihama, 2000).

L'ordre décroissant d'affinité des facteurs σ pour l'apoenzyme est : σ^{70} , σ^{54} , σ^{28} , σ^{32} , σ^{FecI} , σ^{24} et σ^S . De plus, le nombre de molécules des différents facteurs σ en condition de croissance exponentielle a été déterminé grâce à une analyse par western blot quantitatifs (Jishage et Ishihama, 1995; Jishage *et al.*, 1996; Maeda *et al.*, 2000b). Les trois facteurs majoritaires sont σ^{70} (700), σ^{54} (110) et σ^{28} (350), alors que les autres facteurs sont très peu représentés. Sur les 2000 apoenzymes présentes dans la cellule, 1300 sont impliquées dans l'élongation de la transcription. Il reste donc 700 ARN pol libres disponibles pour la fixation d'un facteur σ (Ishihama, 2000). Comme le facteur σ^{70} a le plus d'affinité pour l'apoenzyme et que son taux est le plus élevé dans la cellule en phase exponentielle de croissance, le taux d'holoenzyme $E\sigma^{70}$ est de 78 % (550 molécules), alors qu'il est de 8 % (55 molécules) pour $E\sigma^{54}$ et 14 % (95 molécules) pour $E\sigma^{28}$ (Maeda *et al.*, 2000a; Ishihama, 2000).

Au cours du passage phase exponentielle/phase stationnaire de croissance, la cellule s'adapte à des conditions de croissance défavorables. Comme décrit précédemment, une partie de cette adaptation est réalisée par la modification de la stabilité du facteur σ^S . En effet, la concentration de ce facteur augmente en phase stationnaire de croissance. Dans ce cas, 230 molécules de σ^S sont présentes dans la cellule et s'associent à l'apoenzyme pour permettre la transcription des gènes sous son contrôle (Maeda *et al.*, 2000a). Cependant, le facteur σ^{70} reste majoritaire au cours de la phase stationnaire malgré l'activité d'un facteur antisigma Rsd

(regulation of sigma D) (Jishage et Ishihama, 1998; Jishage et Ishihama, 1999). Rsd se fixe au niveau du domaine σ^4 du facteur σ^{70} notamment au niveau du motif HTH permettant la reconnaissance de la boîte -35. Ce facteur empêche la fixation de σ^{70} sur l'apoenzyme et bloque la reconnaissance des promoteurs reconnus par $E\sigma^{70}$ (mis à part les promoteurs étendus) (Jishage *et al.*, 2001).

Les facteurs σ^{32} , σ^{FecI} et σ^{E} ne sont présents que dans des conditions particulières de croissance comme un choc thermique ou la présence de citrate de fer. Il est à noter que dans un grand nombre de conditions de stress, σ^{70} est toujours présent de façon majoritaire (Ishihama, 2000).

II.2) Implication d'éléments en *cis* dans la modulation de la force d'un promoteur reconnu par l'holoenzyme $E\sigma^{70}$.

L'ARN pol associée au facteur σ^{70} doit sélectionner certains promoteurs parmi les multiples promoteurs présents sur le chromosome. Le premier niveau de régulation est la reconnaissance spécifique de ces promoteurs par l'holoenzyme, qui comme nous l'avons déjà mentionné est basée sur l'interaction des régions 2.4 et 4.2 du facteur σ^{70} et de la séquence des boîtes -10 et -35 (cf. paragraphe I.2.1.1).

D'autres éléments de séquences ADN peuvent avoir une influence sur cette reconnaissance notamment la séquence de la région interboîte, la séquence des régions présentes en amont de la boîte -35 et en aval de la boîte -10.

Le dernier paramètre intrinsèque influençant la force d'un promoteur est la structure de l'ADN. En effet, la molécule d'ADN peut être courbée ou surenroulée, ce qui peut modifier les interactions entre l'ARN pol et le promoteur.

II.2.1) Homologie de la séquence des boîtes -10 et -35 avec le consensus.

La première étape de l'initiation de la transcription est la reconnaissance des boîtes -10 et -35 par l'ARN pol (équation E1) puis la formation du complexe RPo (constante k_3). L'homologie des séquences -10 et -35 avec le consensus reconnu par l'holoenzyme $E\sigma^{70}$ définit l'affinité de l'ARN pol pour l'ADN du promoteur (Knaus et Bujard, 1988). En effet, plus la séquence est proche du consensus, plus la fixation de la polymérase sur les boîtes -10 et -35 est forte (Mulligan *et al.*, 1984; Szoke *et al.*, 1987). Néanmoins, lorsque la séquence d'un promoteur

est identique à la séquence consensus établie par comparaison des promoteurs reconnus par $E\sigma^{70}$, l'étape de fixation de l'ARN pol est améliorée mais l'holoenzyme reste alors bloquée sur le promoteur indiquant que l'étape limitante est devenue l'étape d'échappement (Ellinger *et al.*, 1994a).

Des études statistiques sur les promoteurs reconnus par $E\sigma^{70}$ chez *E. coli* montrent que quasiment aucun promoteur ne possède douze bases homologues au consensus (Deuschle *et al.*, 1986; Mitchell *et al.*, 2003). En effet, la majorité des promoteurs ne présente que 4 résidus sur 6 identiques au consensus de la boîte -10 et 4 résidus sur 6 identiques au consensus de la boîte -35 (Mitchell *et al.*, 2003).

L'influence des séquences des boîtes -10 et -35 est propre à chaque promoteur. En effet, suivant les promoteurs, l'étape limitante de la réaction d'initiation de la transcription est différente. Si l'étape limitante est la formation du complexe RPo, des mutations qui conduisent à diminuer l'efficacité de la liaison de l'ARN pol, affectent aussi la formation du complexe ouvert et la force globale du promoteur. Au contraire, si l'étape limitante est l'échappement de l'ARN pol du promoteur, une mutation dans les séquences -10 et/ou -35 pourra n'avoir que peu d'influence sur la force du promoteur du moment que l'étape limitante n'est pas modifiée. Ainsi, dans le cas du promoteur prmT20 (construit à partir du promoteur λ Prm), avec des boîtes -10 et -35 identiques à la séquence consensus des promoteurs reconnus par $E\sigma^{70}$ (Vo *et al.*, 2003b). Le taux de transcription abortive est très élevé, indiquant que l'étape limitante de l'initiation est l'échappement de l'ARN pol. Des mutations au niveau des boîtes -10 et -35 modifient peu l'initiation de la transcription (Vo *et al.*, 2003b).

L'efficacité de l'initiation de la transcription dépend donc d'un équilibre entre fixation efficace de l'ARN pol sur la séquence promotrice et échappement de l'holoenzyme.

En conclusion, la force d'un promoteur ne dépend pas uniquement de la séquence des boîtes -10 et -35, d'autres paramètres influencent leur efficacité, notamment les séquences autour de ces boîtes.

II.2.2) Influence des séquences autour des boîtes -10 et -35.

D'après le mécanisme d'initiation de la transcription pour l'holoenzyme $E\sigma^{70}$, l'ensemble des séquences des promoteurs joue un rôle dans l'initiation de la transcription (Fig. 8). En effet, les séquences en amont de la boîte -35 ont une influence sur la liaison de l'ARN pol au promoteur, ces contacts étant parmi les premiers aux cours de l'initiation. La séquence de la

région interboîte possède un rôle dans le positionnement spatial des boîtes -10 et -35 et donc influence plusieurs paramètres de la réaction. Enfin, les séquences en aval du site d'initiation de la transcription qui seront transcrites en fin de réaction, sont susceptibles d'intervenir au niveau de la formation des premières étapes d'enchaînement des nucléotides et donc de la transcription abortive.

II.2.2.1) Séquence en amont de la boîte -35.

La séquence de la région d'ADN immédiatement en amont de la boîte -35 a une forte influence sur l'efficacité de la transcription. En effet, la plupart des sites de fixation des facteurs protéiques qui permettent l'activation de la transcription se situent dans cette région. Mais, de manière générale, la séquence située en amont de la boîte -35 intervient dans le mécanisme de base de l'initiation de la transcription.

Pour certains promoteurs tels que les promoteurs des opérons ribosomiques, la présence de régions riches en A et T en amont de la boîte -35 augmente l'efficacité d'initiation de la transcription d'un facteur compris entre 20 et 50 fois (Hirvonen *et al.*, 2001). Dans le cas du promoteur *rrnB* P1, cette séquence appelée élément UP s'étend des positions -40 à -60 par rapport au site d'initiation de la transcription (pour revue : Gourse *et al.*, 2000). Elle peut être transférée dans d'autres promoteurs et activer l'initiation de la transcription. Cette séquence permet la fixation spécifique du domaine C-terminal de la sous-unité α de l'ARN pol (α CTD) (Ross *et al.*, 1993), ce qui augmente l'affinité de l'ARN pol pour le promoteur mais aussi influence d'autres étapes de l'initiation de la transcription (Rao *et al.*, 1994; Strainic *et al.*, 1998). L' α -CTD se fixerait au niveau du petit sillon de l'ADN (Ross *et al.*, 2001; Benoff *et al.*, 2002).

La séquence consensus de cet élément qui a été déterminée par une méthode de sélection *in vitro* basée sur la fixation de l'ARN pol sur des séquences dégénérées est : 5' NNAAAWWTWTTTTNNNAAANN 3' (W= A ou T, N= A, C, G ou T) (Estrem *et al.*, 1998). D'autres résultats montrent que les éléments UP sont composés de deux séquences différentes : une séquence proximale et une séquence distale. Les domaines C-terminaux de chaque sous-unité α de l'ARN pol semblent interagir de façon indépendante sur ces deux séquences (Estrem *et al.*, 1999; Estrem *et al.*, 1998). La fixation de ces deux sous-unités pourrait contribuer de façon indépendante ou additive à l'activation de la transcription (Gourse *et al.*, 2000). Des travaux récents montrent qu'une interaction fonctionnelle entre le

domaine α CTD fixé au site proximal et le domaine $\sigma 4$ fixé au niveau de la boîte -35 permet d'augmenter l'affinité de l'ARN pol pour certains promoteurs (Ross *et al.*, 2003).

La présence d'un élément UP a été détectée au sein de beaucoup d'autres promoteurs (Estrem *et al.*, 1999) notamment le promoteur étendu *galP1* (Busby *et al.*, 1987; Burns *et al.*, 1999).

En accord avec les notions décrites au paragraphe II.2.1, l'augmentation de l'affinité de l'ARN pol pour le promoteur, due à la présence de l'élément UP peut nuire à l'efficacité du promoteur si l'étape limitante est l'échappement de l'holoenzyme (Ellinger *et al.*, 1994a). De plus, la présence d'un élément UP au niveau du promoteur *malT* réduit l'efficacité de ce promoteur par la formation des complexes RPo non productifs (Tagami et Aiba, 1999). Si cet élément UP est décalé de 5 pb ou remplacé par l'élément UP du promoteur *rrnB P1*, l'efficacité du promoteur est renforcée. Les auteurs concluent que la présence d'un élément UP peut entraîner la formation de deux types de complexes RPo productifs ou non productifs en fonction de la position et de la nature de ces éléments UP (Tagami et Aiba, 1999). Cette étude indique que la position de l'élément UP par rapport à la boîte -35 est essentielle à sa fonction dans le mécanisme de l'initiation de la transcription. Cette importance a été confirmée en déplaçant l'élément UP du promoteur *rrnB P1* d'un tour d'hélice. Dans ce cas, l'efficacité du promoteur diminuait fortement (Meng *et al.*, 2001).

Il est à noter que plus la séquence de l'élément UP est proche du consensus, plus la fixation des domaines α -CTD est forte (Ross *et al.*, 1998). Si la séquence est fortement dégénérée (sans élément UP), les domaines α -CTD se fixent sur l'ADN mais de manière non spécifique (Burns *et al.*, 1999). Cette interaction non spécifique n'est pas essentielle à l'initiation de la transcription pour les promoteurs sans éléments UP mais peut tout de même jouer un rôle dans le mécanisme (Burns *et al.*, 1999; Ross *et al.*, 1998).

Les éléments en amont de la boîte -35 constituent donc une zone de contact supplémentaire pour l'ARN pol en plus des contacts entre le facteur σ et les boîtes -10 et -35.

II.2.2.2) Séquence de la région interboîte.

Des études statistiques montrent que pour la majorité des promoteurs reconnus par $E\sigma^{70}$, les boîtes -10 et -35 sont séparées par 17 +/- 1 pb (Harley et Reynolds, 1987; Lisser et Margalit, 1993; Ozoline *et al.*, 1997; Mitchell *et al.*, 2003) et cette distance est très importante pour l'activité du promoteur (Record *et al.*, 1996). En effet, des mutations qui modifient la longueur de la région interboîte, affectent l'initiation de la transcription notamment la

formation des complexes RPo (Warne et deHaseth, 1993). L'importance de cette distance serait liée à la nécessité de bien positionner spatialement les boîtes -10 et -35 par rapport à l'ARN pol, ce qui implique que la taille mais pas forcément la séquence de cette région serait importante pour l'efficacité du promoteur (Auble *et al.*, 1986; Auble et deHaseth, 1988; Ayers *et al.*, 1989; Warne et deHaseth, 1993; DeHaseth et Helmann, 1995). Ainsi, une distance de 16 pb entraîne un positionnement non optimal de la boîte -35 par rapport à la boîte -10, qui se caractérise par une absence de co-linéarité entre les deux boîtes. Une étude statistique montre que les régions interboîtes qui ont des longueurs supérieures à 17 pb ont souvent des séquences riches en enchaînements purines-pyrimidines ou pyrimide-purine alors que les régions interboîtes ayant des longueurs inférieures à 17 pb ont plutôt tendance à avoir des enchaînements purine-purine ou pyrimidine-pyrimidine (Beutel et Record, 1990). Ceci peut être expliqué par le fait que des homodinucléotides présentent des propriétés structurales différentes de celles des hétérodinucléotides. Ainsi, la présence d'homodinucléotides conduit à une torsion (twist) de l'hélice d'ADN plus importante que celle générée par des enchaînements d'hétérodinucléotides. Ces différences de conformation de l'hélice d'ADN de la région interboîte suivant sa composition permettrait de compenser l'angle formé entre les deux hexamères dans le cas de promoteurs ayant une distance interboîte de 16 pb. Il est à noter que dans le cas d'une distance interboîte de 17 pb, aucune préférence en homo- ou hétérodinucléotide n'a été détecté suggérant que la séquence peut être aléatoire (Beutel et Record, 1990).

La région interboîte est courbée au cours du mécanisme d'initiation de la transcription (cf mécanisme, Fig. 8). Comme la séquence de cette région peut influencer sa flexibilité et sa capacité à être courbée, la séquence de la région interboîte peut donc intervenir dans l'efficacité de formation du complexe RPo.

II.2.2.3) Séquences en aval de la boîte -10.

Cette région peut être décomposée en deux éléments de séquence :

- 1) L'élément se situant entre la boîte -10 et le site d'initiation de la transcription (distance boîte -10/(+1)).
- 2) L'élément se situant après le site +1 (jusqu'à +20 dans le cas des complexes d'initiation).

Au cours de la formation des complexes RPo, la région entre la boîte -10 et le site +1 doit être ouverte par l'ARN pol (cf. Fig. 9). La séquence de cette région peut ainsi influencer la

formation et/ou la stabilité des complexes ouverts formés. L'importance de cette séquence a aussi pu être mise en évidence lors de l'étude du promoteur *P_{tyrT}* (Travers, 1980; Pemberton *et al.*, 2000). La séquence de -7 à -1 de ce promoteur, qui est riche en G-C a été appelée discriminateur. Lors de l'initiation de la transcription, la stabilité des complexes ouverts formés à ce promoteur est très faible du fait de la présence de cette région discriminatrice. La séquence riche en G et C pourrait empêcher la formation de complexes RPo stables en bloquant l'ouverture de la double hélice et/ou le déroulement de l'ADN (Pemberton *et al.*, 2000). Les particularités de cette région dans les promoteurs des ARNr pourraient également expliquer la faible stabilité des complexes RPo formés au niveau de ces promoteurs. La région discriminatrice est aussi impliquée dans la probabilité de transcription abortive aux positions +2 à +5 (Vo *et al.*, 2003b).

La séquence de la région d'ADN allant des positions +1 à +20 peut aussi limiter l'efficacité des promoteurs lors de l'étape d'échappement de l'ARN pol (Kammerer *et al.*, 1986). Suivant sa séquence, cette région appelée ITS (internal transcribed sequence) peut en effet affecter la synthèse des produits abortifs en réduisant le taux de transcrits de grande taille (Hsu *et al.*, 1995; Chan et Gross, 2001; Hsu *et al.*, 2003). Plusieurs interprétations sont possibles pour expliquer cet effet. La séquence ITS pourrait modifier la stabilité de l'hybride ADN-ARN, ce qui pourrait bloquer les étapes d'échappement. Une autre explication serait que le facteur σ interagisse avec cette séquence ITS, ce qui limiterait le départ de l'ARN pol. Chan *et al.* (2001) proposent que l'effet de la séquence ITS sur l'initiation de la transcription est médié par des interactions entre la sous-unité σ et cette séquence (Chan et Gross, 2001). De telles pauses au niveau de la séquence ITS ont été détectées pour le promoteur *lacUV5* dans lequel le facteur σ^{70} interagit avec une séquence de la région ITS ressemblant à une boîte -10 (Brodolin *et al.*, 2004).

II.2.3) Influence de la structure de l'ADN.

La séquence de l'ADN n'est pas seulement importante pour la reconnaissance par des protéines mais définit aussi des propriétés structurales, telles que la courbure de l'axe de la double hélice d'ADN, qui joue un rôle important dans tous les mécanismes mettant en œuvre l'ADN notamment la réplication et la transcription. En effet, les courbures naturellement présentes au niveau des promoteurs influencent le mécanisme d'initiation de la transcription.

Le surenroulement de l'ADN circulaire est une autre caractéristique structurale de cette molécule. En effet, le taux de surenroulement peut modifier la structure de la double hélice mais aussi changer son énergie en variant la tension de torsion au sein de l'hélice. L'initiation de la transcription peut être influencée par le degré de surenroulement de l'ADN chromosomique.

II.2.3.1) La courbure de l'ADN.

Comme nous l'avons décrit précédemment, dans le modèle 3D de l'ARN pol en association avec un ADN en fourche, la double hélice est courbée par l'holoenzyme au cours de la formation des complexes d'initiation de la transcription. Des courbures intrinsèques de l'ADN orientées dans la bonne direction pourraient donc favoriser la formation de ces complexes. Dans un premier temps, nous allons décrire les séquences nucléotidiques qui confèrent une courbure intrinsèque de la double hélice d'ADN puis nous décrirons l'effet de ces courbures sur l'initiation de la transcription.

II.2.3.1.1) La formation des courbures au sein de l'ADN.

Des enchaînements de résidus A d'une longueur de 4 à 6 nucléotides en phase avec le pas de l'hélice permettent la formation d'une courbure de l'ADN (pour revue : Perez-Martin *et al.*, 1994; Perez-Martin et de Lorenzo, 1997) (Fig. 16). La présence d'une courbure dans un fragment d'ADN peut être détectée du fait qu'il présente une migration plus lente lors d'électrophorèse en gel par rapport à un fragment de longueur égale parfaitement linéaire (Wu et Crothers, 1984).

Pour expliquer l'induction d'une courbure par un enchaînement de 4 à 6 résidus A au sein d'une double hélice, l'équipe de Crothers a réalisé un grand nombre d'études biochimiques sur de petites molécules (Koo et Crothers, 1988; Koo *et al.*, 1990; Koo *et al.*, 1986; Wu et Crothers, 1984) (pour revue : Crothers *et al.*, 1990). Les auteurs ont proposé un modèle dit de jonction indiquant que la courbure est liée à la présence de distorsions angulaires locales entre les paires de base situées aux jonctions entre l'enchaînement de résidus A et la séquence qui suit (Fig. 16 C) (Crothers *et al.*, 1990). Ces distorsions induisent la compression du petit sillon de l'ADN au centre de l'enchaînement de résidus A donnant ainsi la direction de la courbure (Koo *et al.*, 1986; Crothers *et al.*, 1990). De plus, les auteurs ont montré que les jonctions

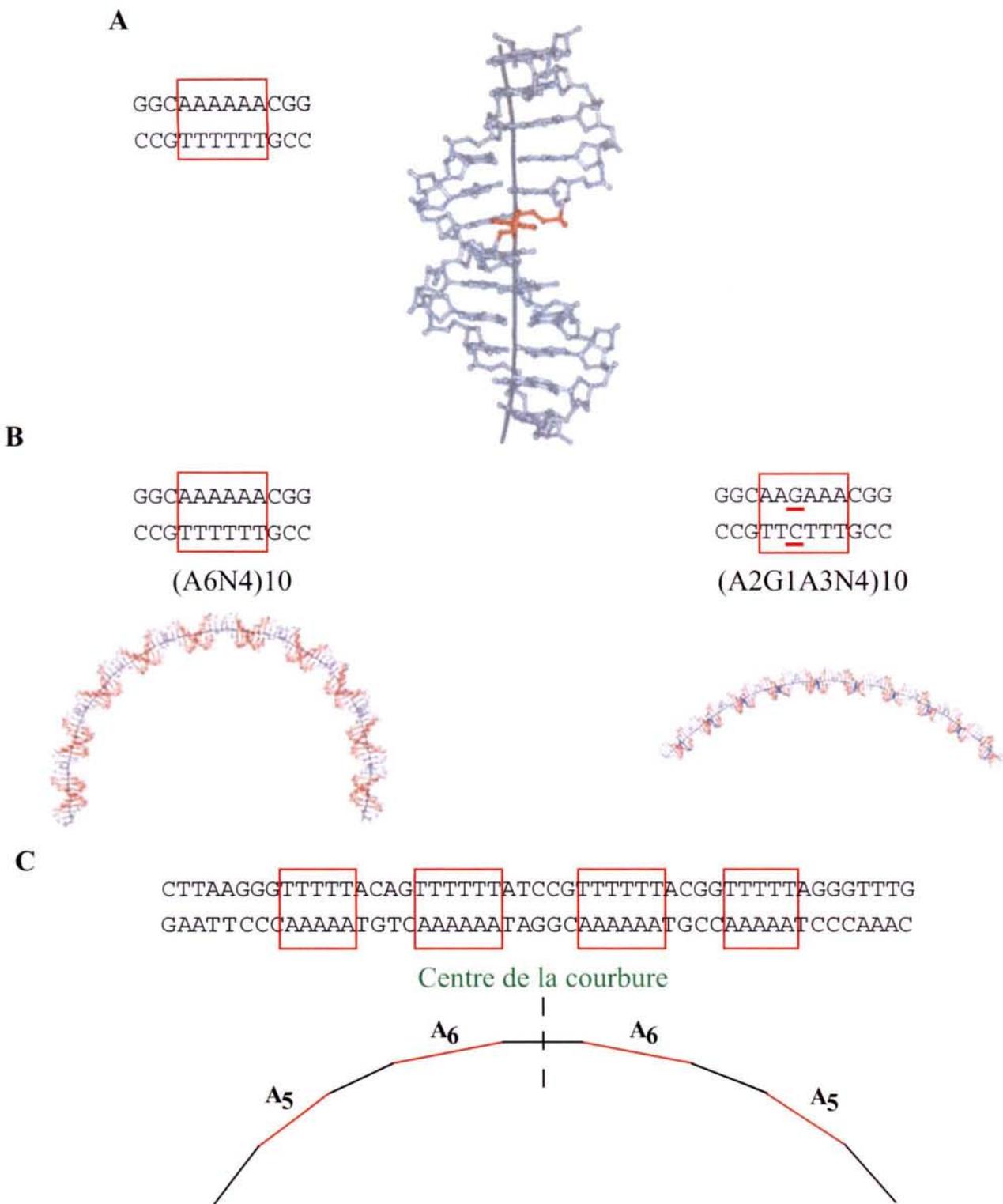


Figure 16 : Formation d'une courbure au sein d'une molécule d'ADN.

A) Structure en solution du dodécamère d'ADN double brin contenant un enchaînement de six résidus A (MacDonald *et al.*, 2001)

B) Structure du dodécamère multimérisé en phase avec le pas de l'hélice (MacDonald *et al.*, 2001).

C) Modèle dit de jonction pour la formation de courbure de l'ADN (Wu et Crothers, 1984).

situées en 5' ou en 3' par rapport à l'enchaînement de résidus A induisaient la formation de courbures différentes. L'amplitude de la courbure est plus forte au niveau de la jonction 3' (Koo *et al.*, 1986; Crothers *et al.*, 1990). Par ailleurs, la séquence flanquant l'enchaînement de résidus A a une importance dans la génération d'une courbure. En effet, les plus fortes courbures ont été observées avec un résidu T en 3' et un nucléotide C en 5' (Koo *et al.*, 1986). Si deux enchaînements de résidus A sont en phase avec le pas de l'hélice (c'est-à-dire séparé par une distance de 10,5 paires de base), les deux courbures s'additionnent. Plus le nombre d'enchaînements de résidus A en phase avec le pas de l'hélice est grand, plus la courbure globale de l'ADN est de grande ampleur (Fig. 16 B et C). C'est ce type de courbure qui est détecté expérimentalement par électrophorèse.

La technique de RMN basée sur le couplage dipolaire a été utilisée pour déterminer la structure de segments d'ADN de 10 ou 12 nucléotides contenant un enchaînement de 4 ou 6 résidus A (Fig. 16 A et B) (MacDonald *et al.*, 2001; Barbic *et al.*, 2003). Cette technique a permis de mettre en évidence que la génération d'une courbure au niveau d'un enchaînement de A dépendait plutôt de la formation de distorsions au niveau des jonctions, ce qui confirme le modèle de Crothers (Fig. 16 A). Ces structures montrent une courbure de 19° décomposée en une courbure de 14° présente au niveau des jonctions à l'extrémité de l'enchaînement de A (4° au niveau 5' et 10° au niveau 3') additionnée d'une courbure de 5° répartie au niveau des paires de base A-T (MacDonald *et al.*, 2001). Ces structures montrent par ailleurs la compression du petit sillon de l'ADN au niveau de l'enchaînement de A. La figure 16 B montre que lorsque 10 séquences de 12 nucléotides sont placées en phase avec le pas de l'hélice, une courbure macroscopique de 180° est générée confirmant les observations biochimiques (MacDonald *et al.*, 2001). Par contre, lorsque l'enchaînement de A est interrompu par un G, la courbure macroscopique est affectée (et est réduite à 90°). Ces résultats seraient expliqués par le modèle de Crothers et par la version améliorée par Barbic *et al.*, (2003), qui ont proposé un modèle dit de courbure délocalisée, qui combine le modèle de jonction et le modèle du plus proche voisin. Ce modèle indique que la courbure est une somme de déformations en phase aussi bien au niveau des jonctions que de l'enchaînement de A (Barbic *et al.*, 2003).

Plusieurs explications ont été données pour le retard de migration de fragments d'ADN courbés lors d'électrophorèse en gel.

Une première explication est que le retard serait dû à une friction plus grande des molécules d'ADN courbé lors du passage dans les pores d'un gel d'acrylamide (MacDonald *et al.*,

2001). En plus, l'ADN courbé aurait une flexibilité plus limitée au cours de la migration, ce qui limiterait son déplacement dans les pores du gel.

Un phénomène important à noter est que le retard de migration maximal est obtenu lorsque la courbure est centrée au niveau du fragment d'ADN alors que ce retard est plus faible si la courbure est plus proche d'une extrémité (Wu et Crothers, 1984).

II.2.3.1.2) Influence de la présence de courbure sur l'efficacité de l'initiation de la transcription.

L'étude de l'effet des courbures de l'ADN sur l'initiation de la transcription a été abordée par deux approches différentes : soit par analyse de l'effet de courbures présentes naturellement dans les régions promoteurs, soit par génération de courbures par mutagenèse de promoteurs. Une étude statistique basée sur la recherche de séries de résidus A au niveau des promoteurs d'*E. coli* a montré la présence fréquente de courbures intrinsèques potentielles au niveau de différentes parties des séquences promotrices, notamment en amont de la boîte -35 (Ozoline *et al.*, 1997). D'autres études indiquent que des courbures présentes au niveau des boîtes -10 et -35 ont aussi une influence sur l'efficacité de l'initiation de la transcription.

L'efficacité de l'initiation de la transcription est affectée aussi bien de façon négative que positive par la présence d'une courbure intrinsèque de l'ADN. L'orientation de ces courbures peut en effet faciliter ou, au contraire, inhiber la fixation de l'ARN pol sur une séquence promoteur (Fig. 17). De la même manière, sachant qu'au cours du mécanisme d'initiation de la transcription la molécule d'ADN s'enroule autour de l'ARN pol, une courbure orientée dans la mauvaise direction pourrait influencer la fixation de l'ARN pol mais aussi les étapes ultérieures du mécanisme d'initiation notamment la formation des complexes ouverts (Perez-Martin *et al.*, 1994) (cf. Fig. 8).

II.2.3.1.2.1) Présence d'une courbure entre les boîtes -10 et -35.

Plusieurs études ont montré l'importance des courbures au niveau de la région interboîte. La première étude concerne une série de promoteurs synthétiques dont les boîtes -10 et -35 possédaient 12 bases identiques à celles de la séquence consensus reconnue par $E\sigma^{70}$ (Lozinski *et al.*, 1989). La région interboîte de ces promoteurs a été modifiée et des séries de A ont été introduites. Les auteurs ont ensuite vérifié par électrophorèse la présence de

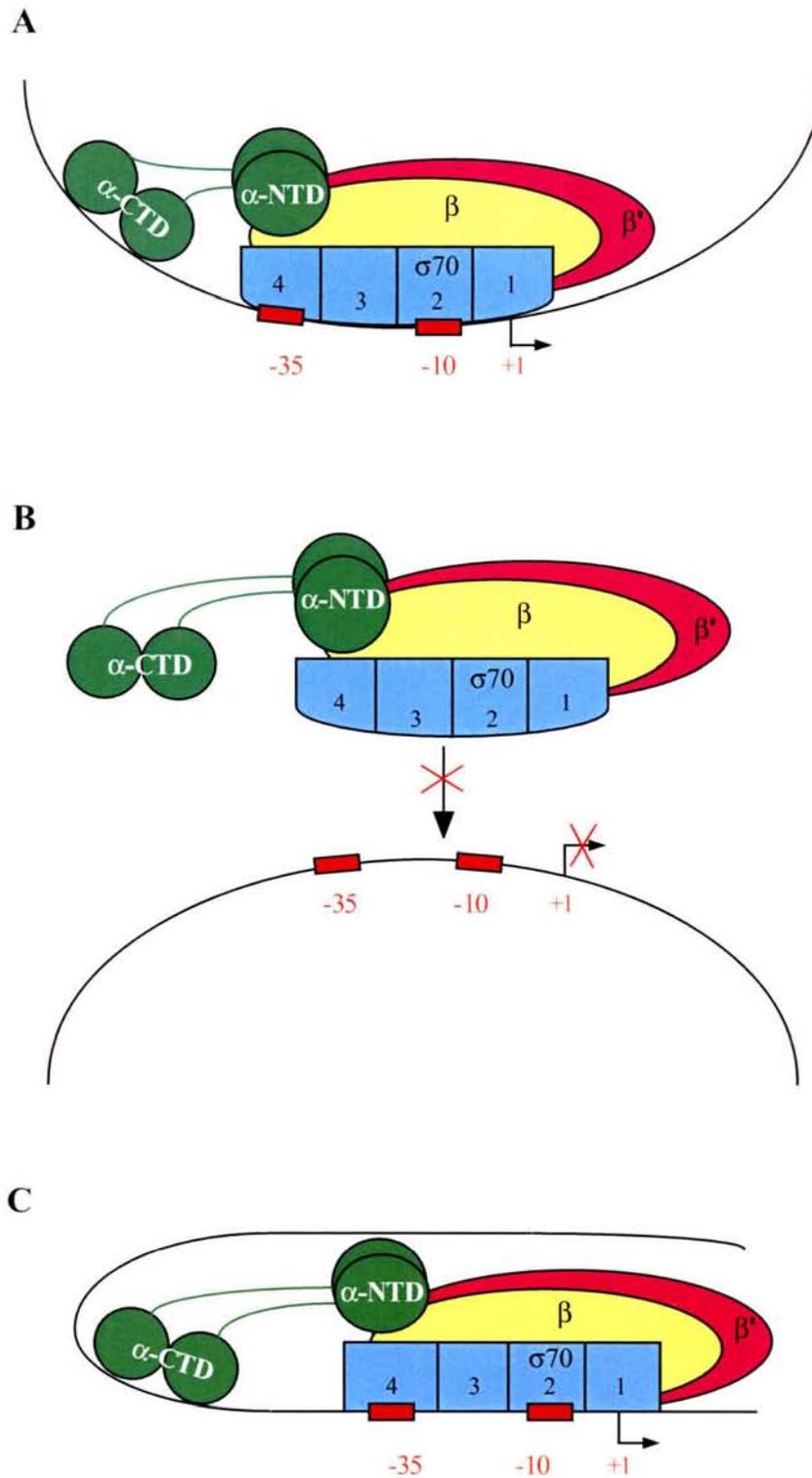


Figure 17 : La présence de courbures de l'ADN au sein d'un promoteur influence l'initiation de la transcription (d'après Perez-Martin *et al.*, 1994).

A) Effet activateur d'une courbure proximale sur l'initiation de la transcription.

B) Si la courbure n'est pas orientée de façon optimale, les interactions de l'ARN pol avec les boîtes -10 et -35 sont défavorisées. Il y a donc inhibition de l'initiation de la transcription.

C) Effet d'une courbure distale sur l'initiation de la transcription. La courbure distale pourrait favoriser l'enroulement de l'ADN autour de l'ARN pol.

courbures au niveau des fragments d'ADN. Les résultats indiquent que l'orientation de la courbure est importante pour l'efficacité de la transcription *in vivo* au niveau de ces promoteurs. Cette même équipe a montré que *in vitro*, l'orientation de cette courbure était importante pour la formation des complexes ouverts (Lozinski *et al.*, 1991). Suivant le positionnement des domaines $\sigma 2$ et $\sigma 4$ de l'ARN pol par rapport à l'ADN, une courbure concave ou convexe peut avoir un effet défavorable sur la fixation de l'holoenzyme mais aussi sur la formation et la stabilité des complexes RPo.

L'étude de l'effet de la courbure de l'ADN a aussi été réalisée sur promoteur *galP1* qui fait partie de la famille des promoteurs étendus (Bown *et al.*, 1997). Au cours de cette étude, Busby *et al.*, (1987) ont muté la région interboîte et créé un enchaînement de résidus A en amont de la boîte -10. L'initiation de la transcription au niveau de ce promoteur est activée, ce qui indique que ce type d'enchaînement de résidus A, qui est connu pour créer une courbure de l'ADN, a une influence sur l'activité du promoteur.

Dans certains cas, la divergence de la séquence des boîtes -10 et -35 par rapport au consensus peut être compensée par la présence d'une courbure orientée dans la bonne direction (Collis *et al.*, 1989). Par contre, une courbure provoquant un mauvais placement des interactions ARN pol-ADN entraîne une diminution de l'activité d'un promoteur (Collis *et al.*, 1989).

Une étude réalisée sur le promoteur synthétique Pa a permis de montrer clairement l'importance de l'orientation de la courbure sur l'efficacité de l'initiation de la transcription (Kolasa *et al.*, 2003). Les auteurs ont testé les effets de la génération d'enchaînements de résidus A dans la région interboîte sur la cinétique de formation des complexes ouverts *in vitro* avec l'holoenzyme $E\sigma^{70}$. Les résultats obtenus montrent que lorsque les enchaînements de résidus A se situent sur le brin non matrice, l'effet négatif sur la transcription est fort. Par contre, lorsque ces deux enchaînements en phase sont placés sur l'autre brin, ils modifient peu la cinétique de formation des complexes ouverts. Les auteurs proposent que les enchaînements de résidus A qui sont situés au niveau des régions courbées formés lors du mécanisme d'initiation de la transcription par l'ARN pol (8° dans la région -25 et 37° dans la région -16), pourraient modifier la liaison de l'ARN pol au promoteur. En effet, la mauvaise orientation de ces courbures serait responsable de la diminution de la formation des complexes ouverts pour ce type de promoteur (Kolasa *et al.*, 2003). Cet effet est aussi observé lorsque la courbure chevauche la séquence de la boîte -35 (Lozinski et Wierchowski, 1996) (Kolasa *et al.*, 2002).

II.2.3.1.2.2) Présence d'une courbure en amont de la boîte -35.

Une proportion importante des promoteurs d'*E. coli* possède une région riche en A en amont de la boîte -35 formant une courbure de l'ADN, dont le centre est situé aux positions -44 et/ou -54 par rapport au site +1 d'initiation de la transcription (Galas *et al.*, 1985; Ozoline *et al.*, 1997). Dans un grand nombre de cas, la présence de ce type de structure de l'ADN permet d'augmenter l'efficacité de la transcription *in vitro* et *in vivo* mais des exemples montrent que pour certains promoteurs, les régions courbées diminuent l'efficacité de la transcription (pour revue : Perez-Martin et Espinosa, 1994).

En amont des promoteurs, la position des courbures peut varier des positions -40 pour le promoteur des opérons ribosomiques d'*E. coli* ou certains promoteurs de phages (courbure proximale) jusqu'à -235 pour le promoteur d'un gène de streptocoque (courbure distale) (Perez-Martin *et al.*, 1994). En général, une seule courbure est présente en amont de la séquence de la boîte -35 mais des exemples montrent que plusieurs courbures formées d'enchaînements de résidus A en phase peuvent se situer en amont de la boîte -35 et influencer l'efficacité d'un promoteur de façon différente (Bauer *et al.*, 1988; Hsu *et al.*, 1991; Perez-Martin *et al.*, 1994). Dans le cas du promoteur *argT*, un enchaînement de résidus A proximal (-60 à -38) augmenterait l'affinité de l'ARN pol pour le promoteur, tandis que l'enchaînement distal (-60 à -130) pourrait favoriser la formation du complexe RPo (Hsu *et al.*, 1991; Perez-Martin *et al.*, 1994). Les courbures les plus distales pourraient avoir un effet sur les étapes tardives du mécanisme de l'initiation de la transcription alors que les premières étapes seraient influencées par les éléments proximaux (Perez-Martin *et al.*, 1994).

La position du centre des courbures amont par rapport au site de fixation de l'ARN pol a un rôle déterminant dans l'effet activateur (Perez-Martin *et al.*, 1994). En effet, lorsque le centre de la courbure est déplacé d'un ou deux tours d'hélice par rapport au promoteur, l'effet activateur de la région courbée est maintenu alors que cet effet disparaît si la région courbée est déplacée d'un demi tour d'hélice (McAllister et Achberger, 1989; Newlands *et al.*, 1992; Perez-Martin et Espinosa, 1994).

Nous avons déjà mentionné l'existence des éléments UP situés en amont de la boîte -35 qui sont des séquences riches en A et T et sont présents en amont d'un grand nombre de promoteurs (pour revue : Gourse *et al.*, 2000). Ces éléments qui peuvent fixer la sous-unité α de l'ARN pol permettent d'activer la transcription. La séquence de ces éléments UP est telle que ces derniers pourraient générer des courbures de l'ADN. Il a été proposé que les

courbures intrinsèques et les éléments UP fonctionnent de la même manière, c'est-à-dire en fixant la sous-unité α de l'ARN pol (α -CTD). La présence de cette interaction a été démontrée en réalisant une étude par empreinte à la DNase I d'un complexe formé entre un mutant de l'ARN pol délété du domaine α -CTD et le promoteur *rrnB* P1 pour lequel l'élément UP avait été remplacé par une région courbée (Aiyar *et al.*, 1998). L'absence de protection de la région en amont de la boîte -35 s'accompagnait d'une diminution de l'activité du promoteur *in vivo* et *in vitro*, confirmant le rôle important de l'interaction entre l' α -CTD et la séquence courbée. Il faut noter que tous les éléments UP ne sont pas des éléments générant une courbure de l'ADN (Gourse *et al.*, 2000).

L'effet des courbures proximales sur l'initiation de la transcription serait donc médié par le domaine α -CTD de l'ARN pol, alors que l'effet des courbures distales serait lié à l'enroulement de la molécule d'ADN autour de l'ARN pol au cours du mécanisme de l'initiation de la transcription. Cet enroulement pourrait favoriser l'ouverture de la double hélice (Coulombe et Burton, 1999; Travers et Muskhelishvili, 1998) comme par exemple au niveau du promoteur *lacUV5* (Amouyal et Buc, 1987) mais aussi au niveau du promoteur *tyrT*, ceci en facilitant le déroulement de l'ADN limité par la région discriminatrice (Pemberton *et al.*, 2002) (Fig. 17). Pour certains promoteurs de phages, l'efficacité est liée à la présence d'une courbure distale qui pourrait alors faciliter l'échappement de l'ARN pol (Perez-Martin *et al.*, 1994).

La présence d'une courbure en amont de la boîte -35 peut entraîner une diminution de l'efficacité d'un promoteur. En effet, dans le cas d'un promoteur limité dans l'étape d'échappement, le renforcement des interactions entre l'ARN pol et la séquence du promoteur (en ajoutant un contact supplémentaire entre l' α -CTD et la région courbée) va encore diminuer l'échappement et ainsi l'efficacité de l'initiation de la transcription (Ellinger *et al.*, 1994b).

En conclusion, la courbure de l'ADN est un paramètre très important de l'initiation de la transcription et son influence est présente à toutes les étapes du mécanisme. Suivant la position de la courbure par rapport au site d'initiation de la transcription et son orientation par rapport aux domaines de l'ARN pol qui permettent la fixation à l'ADN, l'effet de cette courbure sur le mécanisme de la transcription pourra être négatif ou positif. Ces effets sont promoteur-dépendant et notamment dépendant de l'étape limitante du mécanisme pour ce promoteur.

II.2.3.2) Le surenroulement de l'ADN.

II.2.3.2.1) Définition du surenroulement.

Le génome d'*E. coli* est circulaire et sous-enlacé, ce qui implique l'apparition de phénomènes de surenroulement lorsque le double brin subit une torsion dans l'espace autour de l'axe de l'hélice. L'ADN extrait d'*E. coli* possède un surenroulement négatif, qui est dû à une torsion exercée dans la direction opposée à celle de l'orientation de la double hélice (Rouvière-Yaniv *et al.*, 1992; Drlica, 1992). En relâchant la torsion naturelle de l'ADN, ce surenroulement modifie la structure de la double hélice en réduisant la rotation par paire de bases (pour revue : Wang et Syvanen, 1992).

Ce taux de surenroulement négatif est contrôlé *in vivo* par la combinaison de plusieurs facteurs, dont des protéines de liaison à l'ADN, la transcription, la réplication et l'activité des topoisomérases (pour revue : Hatfield et Benham, 2002). Le niveau de surenroulement est défini par deux paramètres, d'une part la fixation de protéines notamment les protéines H-NS, IHF et HU et l'ARN pol qui permettent le contrôle d'environ 50 % du surenroulement de l'ADN d'*E. coli* (pour revue : Pettijohn, 1996; Dorman, 2004), d'autre part par l'action des topoisomérases comme l'ADN gyrase (gènes *gyrA* et *gyrB*) et les topoisomérases I (gène *topA*) et IV (gènes *parC* et *parE*) qui contrôlent le reste du surenroulement (Drlica, 1992; Drlica *et al.*, 1992). L'ADN gyrase qui est composée de deux sous-unités GyrA et GyrB, introduit des supertours négatifs par cassure des deux brins d'ADN, ceci en utilisant de l'énergie fournie par l'hydrolyse de l'ATP, alors que l'enzyme Topoisomérase I composée d'une seule sous-unité élimine ces supertours sans consommation d'ATP. L'enzyme Topoisomérase IV, qui est composée de deux sous-unités, peut aussi éliminer des supertours négatifs en utilisant de l'ATP (Zechiedrich *et al.*, 2000).

Un grand nombre d'études montrent que le taux de surenroulement est contrôlé par le niveau énergétique de la cellule. En effet, lorsque le rapport ATP/ADP qui contrôle l'activité de l'ADN gyrase est faible, le taux de surenroulement négatif diminue et au contraire si le rapport ATP/ADP augmente, le taux de surenroulement augmente (Hsieh *et al.*, 1991; Drlica, 1992; Drlica *et al.*, 1992; Hatfield et Benham, 2002). Le rapport ATP/ADP étant lié aux conditions de croissance d'*E. coli*, le taux de surenroulement de l'ADN pourrait servir de signal pour indiquer des changements environnementaux. En effet, un grand nombre de gènes sont régulés par le taux de surenroulement de l'ADN (Hatfield et Benham, 2002).

II.2.3.2.2) Effet du taux de surenroulement sur l'initiation de la transcription.

Par utilisation d'électrophorèse en gels 2D, Steck *et al.*, (1993) ont montré que l'expression d'un grand nombre de protéines est modifiée dans des souches mutées dans l'ADN gyrase (*gyrB*) ou dans la Topoisomérase I (*topA*). Les protéines ont été réparties en trois classes selon qu'elles sont exprimées de manière optimale pour : 1) un taux de surenroulement observé dans les conditions physiologiques normales, 2) une densité de surenroulement négatif élevé et 3) au contraire une densité de surenroulement négatif faible.

Ces différences de taux d'expression en fonction du surenroulement seraient dues à la modification de l'efficacité de l'initiation de la transcription au niveau de certains promoteurs (Pruss et Drlica, 1989; Steck *et al.*, 1993). Des changements dans le taux de surenroulement modifient la structure de l'ADN et comme nous l'avons vu précédemment, la formation des complexes avec l'ARN pol est très sensible à de tels changements.

En particulier, des variations du taux de surenroulement pourraient modifier l'orientation relative dans l'espace des boîtes -10 et -35, limitant ou favorisant les interactions entre les régions $\sigma 2$ et $\sigma 4$ et l'ADN du promoteur suivant le déplacement généré (Wang et Syvanen, 1992). Sachant qu'un taux de surenroulement négatif favorise la détorsion de l'ADN, la distance de la région interboîte peut être modifiée par le surenroulement. Donc, pour des promoteurs ayant une région interboîte sous-optimale de 16 pb, une diminution du taux de surenroulement négatif de l'ADN permet d'augmenter la distance et de placer les séquences -10 et -35 à des positions favorisant la mise en place des interactions avec les domaines $\sigma 2$ et $\sigma 4$. Au contraire, si le taux de surenroulement négatif augmente, les interactions entre le facteur σ et les boîtes -10 et -35 seront défavorisées, ce qui limitera l'initiation de la transcription (pour revue : Wang et Syvanen, 1992; Hatfield et Benham, 2002). Comme attendu, les promoteurs qui possèdent une région interboîte de 17 pb ont tendance à avoir une activité transcriptionnelle optimale pour des taux de surenroulement physiologiques (Steck *et al.*, 1993).

Un autre effet des changements du taux de surenroulement peut être la formation de région d'ADN où la structure en double brin est déstabilisée. Cette déstabilisation pourrait favoriser la formation des complexes RPo, ceci en diminuant l'énergie nécessaire à l'ouverture de la double hélice. L'effet sera tout particulièrement positif dans le cas d'un promoteur où la formation du complexe ouvert est l'étape limitante (Hatfield et Benham, 2002; Revyakin *et*

al., 2004). Des travaux sur le promoteur *tyrT* qui possède une région discriminatrice ont en effet montré qu'un surenroulement négatif favorisait le déroulement de l'hélice et ainsi permettait la formation du complexe RPo (Drew *et al.*, 1985). Le promoteur du gène *hisR*, qui est sensible au surenroulement, constitue un autre exemple. Il possède une région discriminatrice entre les positions -7 à +1. Une mutation dans la région discriminatrice rend le promoteur moins sensible au surenroulement. En effet, le surenroulement négatif modifie l'énergie nécessaire pour ouvrir la double hélice d'ADN et favorise la formation des complexes RPo (Figuroa-Bossi *et al.*, 1998). Au cours de travaux réalisés sur le promoteur étendu *galP1* pour lequel les complexes RPo peuvent se former à basse température (Burns et Minchin, 1994), il a aussi été montré que le surenroulement permet la formation des complexes ouverts à plus basse température, ceci en diminuant l'énergie nécessaire à l'ouverture de la double hélice (Burns et Minchin, 1994).

En étudiant l'effet du surenroulement sur les différentes étapes de l'initiation de la transcription, Lim *et al.*, (2003) ont montré que le taux de surenroulement affectait l'initiation mais aussi l'étape d'échappement de l'ARN pol et même l'étape de terminaison de la transcription (Lim *et al.*, 2003). De plus, ces auteurs ont confirmé le fait que l'influence du surenroulement est dépendante de la séquence du promoteur. En conclusion, des modifications dans le taux de surenroulement de l'ADN peuvent influencer l'initiation de la transcription de façon positive ou négative suivant la séquence du promoteur mais aussi suivant la nature de l'étape limitant l'initiation à ce promoteur.

Un promoteur bactérien est donc composé de différents éléments agissant en *cis* qui permettent de définir son efficacité : boîtes -10 et -35, région interboîte importante par sa taille, sa structure et sa flexibilité, élément UP ou élément courbé en amont de la boîte -35. Par ailleurs, le degré de surenroulement de l'ADN qui peut changer la conformation de ces éléments joue un rôle très important. La combinaison des propriétés de ces différents éléments ADN permet de créer une hiérarchie parmi les séquences promotrices. Les plus efficaces sont celles qui initient de façon prioritaire la transcription des gènes essentiels à la survie de la bactérie. Les promoteurs parmi les plus forts sont par exemple les promoteurs des opérons ribosomiques. Leur transcription en phase active de division cellulaire est essentielle car elle permet la synthèse des ribosomes permettant la traduction des ARNm en protéines. Les séquences des boîtes -10 et -35 de ces promoteurs possèdent 11 bases sur 12 homologues au consensus $E\sigma^{70}$. Un élément UP existe en amont de la boîte -35. Par contre, la séquence de la région interboîte est de 16 pb, indiquant que les promoteurs forts ne possèdent pas forcément

des architectures répondant à tous les critères du consensus. Pour répondre aux changements des conditions environnementales, cette hiérarchie d'efficacité doit être modifiée, ceci afin de permettre la transcription des gènes qui deviennent essentiels dans ces nouvelles conditions. Le changement du taux de surenroulement pourrait être un des paramètres qui permettrait un changement global de l'expression des gènes (Hatfield et Benham, 2002). Un autre moyen est l'utilisation de facteurs agissant *en trans* qui permettent, en fonction de leur taux de production ou de leurs activités, d'optimiser la quantité d'ARN synthétisés et ainsi de modifier la hiérarchie de transcription des gènes en fonction des conditions de croissance de la bactérie.

II.3) Régulation par des facteurs agissant *en trans*.

Sachant que la majorité des promoteurs ont des boîtes -10 et -35 dont la séquence diverge par rapport à celle de la séquence consensus et que les séquences situées autour de ces boîtes ont un rôle local, la sélection par l'ARN pol de familles de promoteurs pour l'initiation de la transcription en fonction des conditions extérieures est souvent effectuée grâce à des facteurs agissant *en trans*. Il peut s'agir de petites molécules, mais surtout de facteurs protéiques se fixant à l'ADN. Ces facteurs peuvent agir de façon positive ou négative au niveau de toutes les étapes du mécanisme d'initiation de la transcription.

II.3.1) Régulation par des molécules de petite taille.

II.3.1.1) Concentration en nucléotide initiateur.

Depuis longtemps, il a été observé que la quantité de ribosomes dans des cellules qui sont en croissance rapide est plus élevée que dans des cellules en croissance lente. Néanmoins, le mécanisme de ce contrôle n'est pas encore décrypté (pour revue : Keener et Nomura, 1996). En effet, des travaux de l'équipe de R. Gourse avait conduit à une explication possible, mais des travaux très récents montrent que l'hypothèse proposée n'était pas valable. Nous décrivons l'ensemble de ces données. La synthèse des ribosomes est limitée par le taux de synthèse des ARNr, qui varie de manière linéaire avec le carré de la vitesse de croissance (Gausing, 1977), ceci du fait que l'initiation de la transcription au niveau du promoteur des opérons ribosomiques *rrnB* P1 est fortement régulée par la vitesse de croissance. Plus la vitesse de croissance est élevée, plus l'activité des promoteurs des opérons ribosomiques est

forte. Il a été montré que la concentration en nucléotide initiateur ATP ou GTP permet de réguler l'activité des promoteurs *rrnB* P1 et *rrnD* P1 (pour revue : Schneider *et al.*, 2003). Pour que l'efficacité du promoteur *rrnB* P1 soit optimale *in vitro*, il faut que la concentration en ATP soit anormalement élevée par rapport à ce qui est observé pour d'autres promoteurs (Gaal *et al.*, 1997). Les complexes RPo formés au niveau du promoteur *rrnB* P1 sont instables, certainement du fait de la présence d'une région discriminatrice. L'idée est qu'une forte concentration en nucléotide initiateur permettrait la stabilisation de ces complexes (Gaal *et al.*, 1997).

L'équipe de R. Gourse a alors proposé que cet effet explique l'existence d'une régulation de ce promoteur *in vivo* en fonction de la vitesse de croissance, car selon les données de cette équipe, la concentration intracellulaire en ATP varie avec la vitesse de croissance (Gaal *et al.*, 1997). Cette équipe en avait déduit que l'effet du nucléotide initiateur *in vivo* est direct. En effet, ces auteurs avaient artificiellement modifié le taux d'ATP dans les cellules en mutant des gènes du métabolisme des purines. Dans ce cas, l'activité *in vivo* du promoteur *rrnB* P1 était modifiée (Schneider *et al.*, 2002; Murray *et al.*, 2003).

Ces résultats ont été sujets à controverse, puisqu'une autre équipe a montré en utilisant une autre technique de mesure de l'ATP, que le taux d'ATP ne varie pas au cours de la croissance (Petersen et Moller, 2000) et récemment, l'équipe de R. Gourse a montré que la variation de concentration en ATP détectée dans leurs expériences antérieures était due à la technique de mesure de l'ATP employée (Schneider et Gourse, 2004). Il existerait donc un autre mécanisme de régulation encore non connu, qui lie le taux d'expression des ARNr à la vitesse de croissance.

Ce type de régulation a été montré pour le promoteur *rrnB* P1 mais pourrait être valable pour tous les promoteurs qui sont limités au niveau de la stabilité des complexes RPo.

II.3.1.2) Concentration en ppGpp.

Le ppGpp ou guanosine 5'-diphosphate 3'-diphosphate est une molécule synthétisée lors d'une carence en acides aminés mais aussi dans un grand nombre de conditions de croissance. Cette molécule peut modifier directement ou indirectement l'expression d'un grand nombre de gènes, ceci afin d'ajuster le métabolisme cellulaire aux conditions de carence. Ce phénomène a été appelé réponse stringente (pour revue : Cashel *et al.*, 1996). L'augmentation de ppGpp entraîne une forte diminution de la synthèse des ARNr, ce qui limite la quantité de ribosomes. En contrepartie, l'expression des opérons de transport et de biosynthèse des acides

aminés est augmentée. La présence de ppGpp ne serait pas impliquée dans le contrôle de l'activité des promoteurs des opérons ribosomiques par la vitesse de croissance car pour des souches ne produisant plus de ppGpp, l'activité du promoteur *rrnB* P1, qui est soumise au contrôle stringent, est toujours dépendante de la vitesse de croissance (Gaal et Gourse, 1990; Bartlett et Gourse, 1994).

Le mécanisme d'action du ppGpp au niveau des promoteurs des ARN stables (opérons ribosomiques et ARNt) a été étudié et les séquences responsables de cet effet ont été mises en évidence. Les séquences impliquées se situent entre les positions -41 à +1, ce qui confirme l'implication du ppGpp dans l'initiation de la transcription (Josaitis *et al.*, 1995). Ces séquences impliquées contiennent une région discriminatrice riche en G et C, située en amont du site +1, qui est certainement responsable de la faible stabilité des complexes RPo (Travers, 1980; Barker *et al.*, 2001b; Barker *et al.*, 2001a; Cashel *et al.*, 1996). Cette région discriminatrice est en effet indispensable à la limitation de l'activité du promoteur *rrnB* P1 par le ppGpp (Josaitis *et al.*, 1995).

L'effet négatif du ppGpp sur l'initiation de la transcription au niveau du promoteur *rrnB* P1 a été étudié *in vitro* et *in vivo*. Cette molécule semble diminuer la stabilité des complexes RPo formés au niveau de ce promoteur *in vitro*. Elle pourrait agir de la même manière au niveau de tous les promoteurs qui sont sa cible (Barker *et al.*, 2001a; Artsimovitch *et al.*, 2004). L'effet du ppGpp ne se ferait donc sentir que sur des promoteurs dont la stabilité des complexes RPo est limitée comme les promoteurs des opérons ribosomiques.

L'idée est que le ppGpp interagisse directement avec l'ARN pol et modifie son site actif, ce qui pourrait affecter la fixation des nucléotides dans ce site (Artsimovitch *et al.*, 2004). Sachant que la fixation des nucléotides initiateurs augmente la stabilité des complexes RPo formés au niveau du promoteur *rrnB* P1 (Gaal *et al.*, 1997), l'effet global serait une diminution de la stabilité de ces complexes et donc une inhibition de la transcription. Il y aurait donc une compétition entre la stabilisation par le nucléotide initiateur et la déstabilisation des complexes ouverts par le ppGpp (Jores et Wagner, 2003). Une autre hypothèse serait que le ppGpp interagisse avec les bases C du brin non matrice dans le site actif de l'ARN pol (Artsimovitch *et al.*, 2004). Cette liaison pourrait modifier des interactions ARN pol-ADN qui stabilisent le complexe RPo, ce qui réduirait ainsi la stabilité de ce complexe (Artsimovitch *et al.*, 2004).

La concentration en nucléotide initiateur et en ppGpp permettrait donc de réguler de façon globale l'expression des gènes. Ces facteurs agissant en *trans* ont un effet sur des promoteurs

(ARNr et ARNt) dont la caractéristique est une faible stabilité des complexes RPo. Ces promoteurs cibles de ces régulations négatives font néanmoins partie des promoteurs les plus efficaces chez les bactéries. L'action des nucléotides permet néanmoins un contrôle strict de leur activité qui du fait du coût énergétique très élevé de la traduction doit être adaptée. Cependant, la majorité des autres gènes d'*E. coli* est surtout régulée par des facteurs de nature protéique, qui sont exprimés et/ou activés dans certaines conditions de croissance. Nous allons décrire certaines de ces régulations.

II.3.2) Régulation par un facteur protéique.

Les facteurs protéiques de transcription permettent une régulation très fine de l'expression des gènes pour que le développement de la bactérie soit optimisé en fonction des conditions environnementales (pour revue : Browning et Busby, 2004).

Le génome d'*E. coli* contient approximativement 300 gènes qui codent des protéines distinctes pouvant se fixer au niveau de régions promotrices pour activer ou réprimer l'initiation de la transcription. Il est estimé qu'environ la moitié de ces gènes codent effectivement des protéines qui ont une fonction de régulateur de l'initiation de la transcription (Madan Babu et Teichmann, 2003b; Perez-Rueda et Collado-Vides, 2000). Ces facteurs de transcription permettent de créer des réseaux de régulation complexes.

La majorité de ces protéines régulatrices se fixent sur l'ADN au niveau de séquences spécifiques, ce qui permet de réguler sélectivement les promoteurs. D'autre part, ces facteurs de transcription peuvent soit être spécifiques d'un facteur σ donné, soit influencer l'initiation de la transcription par des ARN pol associées à différents facteurs σ (Perez-Rueda et Collado-Vides, 2000). Nous ne décrivons que des régulations s'exerçant sur l'holoenzyme $E\sigma^{70}$.

Lorsque les facteurs de transcription se fixent au niveau d'un promoteur, ils peuvent soit être des activateurs de la transcription, soit des répresseurs. Dans les deux cas, ils agissent en activant ou inhibant une ou plusieurs étapes du mécanisme d'initiation de la transcription (pour revue : Browning et Busby, 2004). Certains facteurs agissent seulement comme des activateurs ou des répresseurs, alors que d'autres peuvent être activateurs ou répresseurs suivant le promoteur cible (Perez-Rueda et Collado-Vides, 2000). L'induction d'un effet activateur ou répresseur est souvent lié à la position du site de fixation du facteur de transcription sur l'ADN par rapport aux séquences reconnues par l'ARN pol (Perez-Rueda et Collado-Vides, 2000; Madan Babu et Teichmann, 2003b; Madan Babu et Teichmann, 2003a).

Sept facteurs protéiques sont impliqués dans la régulation d'environ 50% des gènes d'*E. coli* : CRP, ArcA, Fis, IHF, FNR, NarL et Lrp. La majorité des autres facteurs protéiques régulent un seul ou quelques gènes (Martinez-Antonio et Collado-Vides, 2003).

Les facteurs de transcription peuvent être regroupés selon différentes familles, soit sur la base d'homologies de séquence (Perez-Rueda et Collado-Vides, 2000), soit sur la base de la position des contacts qu'ils établissent avec les sous-unités de l'ARN pol (Ishihama, 2000) (Fig. 18). La majorité de ces contacts concernent les sous-unités α et σ , qui sont en effet les cibles des facteurs de transcription. Ces contacts et leur influence sur l'initiation de la transcription, qui ont été démontrés notamment par des expériences de pontages protéines-protéines, des études sur l'expression *in vivo* et sur l'effet de mutations sur l'efficacité de transcription *in vitro* seront décrits plus loin dans ce chapitre.

La fixation des facteurs de transcription sur leurs séquences cibles (opérateurs) est régulée en réponse aux variations des conditions environnementales. Le paramètre le plus souvent utilisé pour réguler le taux d'occupation de l'opérateur est une variation de la concentration du facteur de transcription dans la cellule. De plus, les capacités de fixation sur l'ADN, c'est-à-dire l'affinité des facteurs pour l'ADN cible peuvent être modifiées par des effecteurs notamment la fixation d'un cofacteur dans le cas du complexe CRP-AMPc ou la phosphorylation dans le cas du facteur OmpR. En effet, la fixation de la molécule de cofacteur conduit à des changements de la structure du facteur de transcription. Les capacités de fixation d'un facteur de transcription peuvent aussi être modifiées par sa séquestration au niveau d'une zone de la cellule, ce qui empêche son activité sur le chromosome.

II.3.2.1) Modulation de la fixation des facteurs agissant en *trans* sur les séquences régulatrices.

II.3.2.1.1) Variation de la concentration en facteurs de transcription.

La concentration en facteur de transcription permet de réguler l'expression d'un grand nombre de gènes en fonction des phases de croissance de la bactérie et de la disponibilité en nutriments. Cette concentration est déterminée par le niveau d'expression du facteur mais aussi par son taux de protéolyse dans la cellule (Browning et Busby, 2004). Par exemple, la quantité de la protéine Fis, qui régule l'expression des gènes des opérons ribosomiques et des ARNt, est très élevée en phase exponentielle de croissance et diminue en phase stationnaire

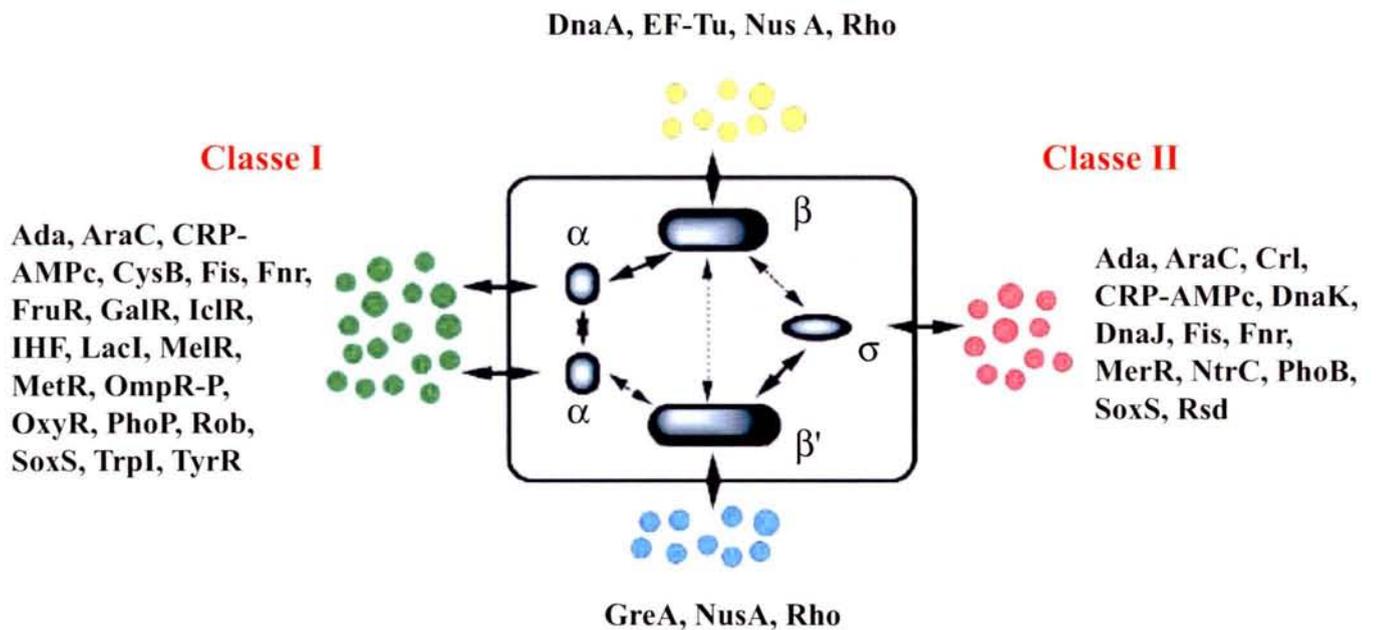


Figure 18 : Classification des facteurs de transcription en fonction des zones de contact qu'ils établissent avec les sous-unités de l'ARN pol (d'après Ishihama, 2000). Les facteurs de transcription de classe I entrent en contact avec la sous-unité α , alors que les facteurs de transcription de classe II entrent en contact avec la sous-unité σ . Les autres facteurs interagissent avec les sous-unités β et β' . Sachant que certains facteurs de transcription ont à la fois des rôles d'activateurs et de répresseurs de la transcription, il n'y a pas de distinction entre ces activités pour la fixation au niveau des sous-unités de l'ARN pol.

de croissance, alors que, dans le cas de la protéine IHF, le taux de cette protéine augmente en phase stationnaire de croissance. Ces différences seraient dues dans les deux cas à des régulations au niveau du promoteur de leurs gènes respectifs, *fis* et *ihf* (Ball *et al.*, 1992; Aviv *et al.*, 1994; Murtin *et al.*, 1998; Walker *et al.*, 1999; Mallik *et al.*, 2004). Les concentrations des douze protéines majeures, qui se fixent à l'ADN, ont été mesurées en phase exponentielle de croissance et en phase stationnaire de croissance (Azam et Ishihama, 1999). C'est en phase exponentielle que la concentration des protéines Fis, HU, IHF, StpA et H-NS est la plus élevée. En phase stationnaire de croissance, les facteurs de transcription majoritairement exprimés sont Dps, IHF et HU (Azam et Ishihama, 1999).

Cette régulation par effet de concentration s'appliquerait surtout aux facteurs de transcription généraux, qui régulent la transcription d'un grand nombre de promoteurs. Lorsque la concentration de ces facteurs augmente dans la cellule, le taux d'occupation des différents opérateurs est modifié suivant leur affinité pour le facteur de transcription.

II.3.2.1.2) Modification par liaison d'un cofacteur : exemple du complexe CRP-AMPc.

Le complexe CRP-AMPc est un des régulateurs de la transcription les plus étudiés chez *E. coli* (pour revue : Kolb *et al.*, 1993a; Busby et Ebright, 1999). Ce complexe régule la transcription d'une centaine de gènes et est impliqué dans les mécanismes de répression catabolique. La protéine CRP (cAMP receptor protein appelée aussi CAP pour catabolic activator protein) est codée par le gène *crp* et semble fortement conservée chez les bactéries Gram⁻. Cette protéine, qui est sous forme d'homodimère n'acquiert sa capacité de fixation à l'ADN et donc de régulation de la transcription qu'en présence d'AMP cyclique (Adénosine 3', 5'-monophosphate cyclique ou AMPc), qui peut se fixer sur chaque monomère (pour revue : Harman, 2001). Chaque monomère est composé de deux domaines (Fig. 19 A). La partie N-terminale (1-133) possède un site de fixation pour l'AMPc et le site de dimérisation (en vert dans la figure 19 A). La partie C-terminale (133-209) possède le site de fixation à l'ADN (motif HTH) (en rouge dans la figure). Ces deux domaines sont reliés par une charnière flexible. En solution, il peut donc exister trois conformères : le dimère CRP, le dimère CRP avec un seul AMPc fixé (CRP-AMPc1) et le dimère CRP avec deux AMPc (CRP-AMPc2) (Heyduk et Lee, 1989). La structure du complexe CRP-AMPc2 est présentée dans la figure 19 B.

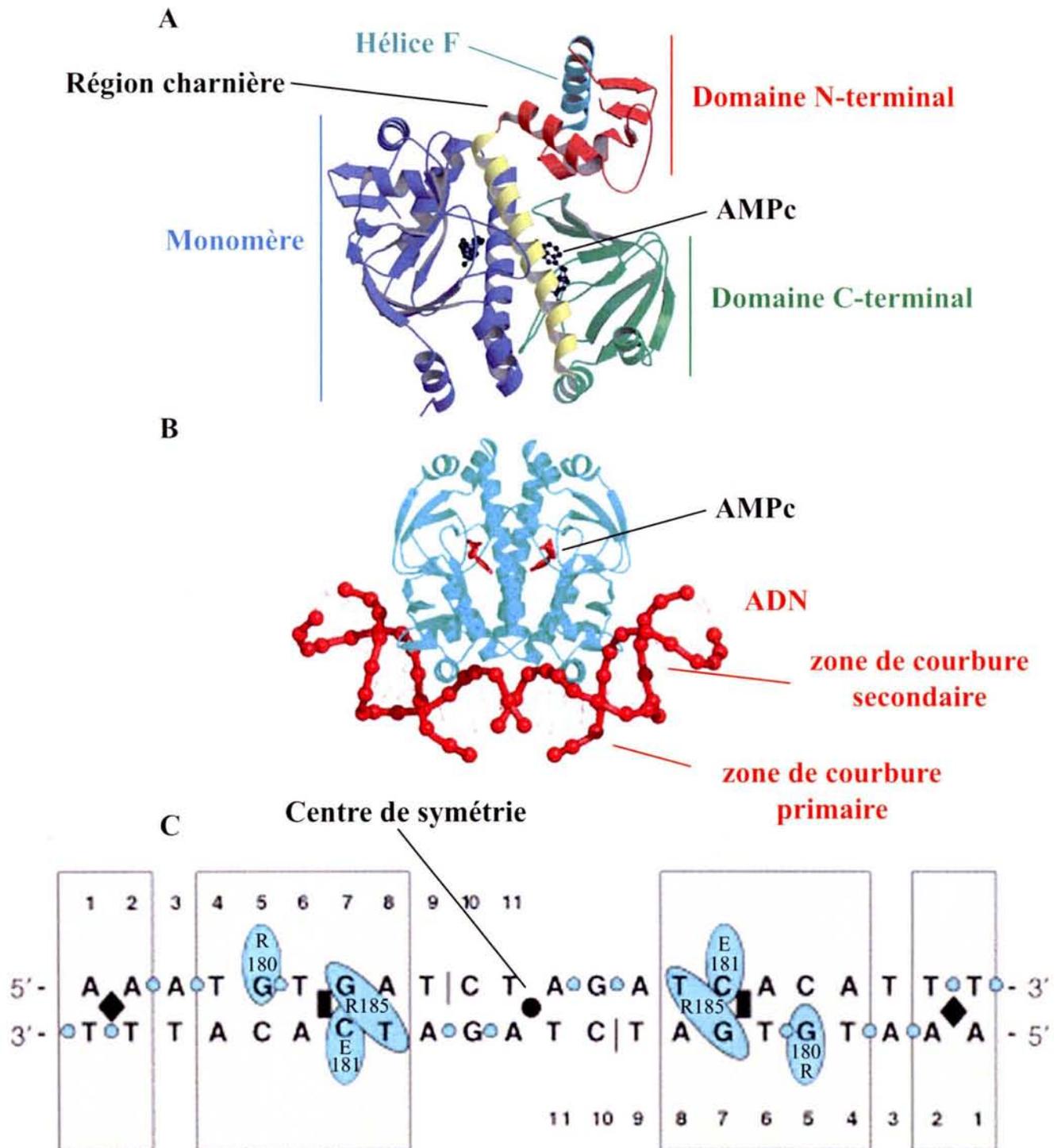


Figure 19 : Bilan des données obtenues à partir des études structurales 3D du complexe CRP-AMPc et de ce complexe en association avec une molécule d'ADN.

A) Structure 3D du complexe CRP-AMPc. Dans une sous-unité, le domaine de fixation de l'AMPc est représenté en vert et le domaine de fixation à l'ADN en rouge. L'hélice C qui est impliquée dans la liaison entre monomères est indiquée en jaune. L'hélice F qui permet la reconnaissance de l'ADN est indiquée. La molécule d'AMPc est représentée en noir. La deuxième sous-unité est représentée en bleu (d'après Dong *et al.*, 2002).

B) Structure 3D du complexe CRP-AMPc associé à une molécule d'ADN (d'après Lawson *et al.*, 2004). La molécule d'ADN est représentée en rouge et les deux sites de courbures sont indiqués. Le complexe CRP-AMPc est représenté en bleu avec l'AMPc en rouge.

C) Bilan des interactions s'établissant entre le complexe CRP-AMPc2 et l'ADN (d'après Lawson *et al.*, 2004). Les boîtes en gris indiquent les résidus de l'ADN en contact avec le complexe. Les rectangles et les losanges noirs indiquent les courbures primaires et secondaires. Le cercle noir indique le centre de symétrie. Les cercles bleus indiquent les positions où un contact est établi entre acides aminés et phosphate et les ovales bleus les positions où des contacts ont lieu entre acides aminés et bases de l'ADN.

Lorsqu'une molécule d'AMPc est déjà fixé, le dimère possède une coopérativité négative pour la fixation d'AMPc dans le deuxième site (pour revue : Harman, 2001). La concentration en AMPc doit donc être élevée pour obtenir la forme CRP-AMPc₂, ce qui ne semble pas être le cas dans une cellule. Le complexe majoritaire *in vivo* serait donc CRP-AMPc₁ (Heyduk et Lee, 1989). De plus, l'affinité pour l'ADN et la spécificité de fixation à l'ADN du complexe CRP-AMPc₁ est plus élevée que celle du complexe CRP-AMPc₂ (Heyduk et Lee, 1990; Heyduk et Lee, 1989; Pyles et Lee, 1996).

Différentes méthodes ont été utilisées pour montrer que la fixation d'une molécule d'AMPc sur le dimère de protéine CRP entraîne des changements de structure de faible amplitude des protéines, qui permettent d'exposer des éléments interagissant avec l'ADN (pour revue : Harman, 2001). Ces modifications structurales entraînent l'expulsion à la surface de l'homodimère de l'hélice F (Baichoo et Heyduk, 1997; Won *et al.*, 2000). Ces changements structuraux créent la conformation qui permet l'interaction spécifique du complexe CRP-AMPc avec l'ADN, mais aussi avec l'ARN pol. Dans le cas de la fixation de deux molécules d'AMPc sur le dimère, les changements structuraux seraient plus importants : le taux d'hélice α diminuerait de 7 %, alors que le taux de feuillet β augmenterait de 7 % (DeGrazia *et al.*, 1990; Tan *et al.*, 1991).

Le site de fixation du complexe CRP-AMPc est composé d'une séquence palindromique de 22 pb. Cette séquence (5' AAATGTGA-N6-TCACATTT 3') a été déterminée par l'alignement de séquences de sites de fixation connus du complexe (Berg et von Hippel, 1988; de Crombrughe *et al.*, 1984). De plus, la structure cristallographique du complexe CRP-AMPc en association avec une molécule d'ADN a permis de préciser les acides aminés de CRP et les résidus de l'ADN impliqués dans l'interaction (Parkinson *et al.*, 1996; Passner *et al.*, 2000; Passner et Steitz, 1997). Un monomère de CRP interagit avec la première séquence répétée, alors que l'autre monomère interagit avec l'autre séquence inversée répétée (Ebright *et al.*, 1989). Il existe donc une symétrie dans les sites de fixation de ce complexe (Fig. 19 C). Les bases qui sont directement en interaction avec les acides aminés du complexe sont indiquées dans la figure 19 C. Elles se situent au niveau des séquences répétées inversées 5' TGTGA 3' aux positions 4 à 8 (Fig. 19 C) et au niveau des deux premières positions 1 et 2 (Fig. 19 C) (Lawson *et al.*, 2004; Ebright *et al.*, 1989; Schultz *et al.*, 1991). Pour chaque monomère, les acides aminés partenaires sont les résidus Arg 180 et Arg 185, ainsi que Glu 181 (Zhang et Ebright, 1990; Parkinson *et al.*, 1996; Ebright *et al.*, 1987). La plupart des

autres contacts indiqués dans la figure 19 C seraient indirects (Schultz *et al.*, 1991; Parkinson *et al.*, 1996).

Des modifications de la séquence de l'opérateur entraînent une diminution de l'affinité du complexe CRP-AMPc. En effet, outre la séquence palindromique qui interagit directement avec les acides aminés de CRP-AMPc, la région centrale notamment la distance séparant les deux séquences palindromiques mais aussi les séquences flanquantes peuvent influencer l'affinité du complexe pour l'ADN (Barber *et al.*, 1993; Pyles *et al.*, 1998a).

Le complexe CRP-AMPc se fixe très fortement sur un ADN ayant la séquence consensus. Les variations de concentration en complexe CRP-AMPc dans la cellule ne permettraient pas une dissociation du complexe si l'opérateur avait exactement la séquence consensus (Gaston *et al.*, 1989), ces promoteurs seraient donc toujours associés au complexe CRP-AMPc et seraient constitutifs. En fait, la plupart des sites de fixation de CRP-AMPc sur le génome d'*E. coli* sont divergents par rapport au consensus, ceci afin de permettre une modulation de la fixation du complexe et donc une régulation de l'initiation de la transcription.

La fixation du complexe CRP-AMPc sur l'opérateur courbe l'ADN d'un angle d'environ 80° (Schultz *et al.*, 1991; Parkinson *et al.*, 1996; Kapanidis *et al.*, 2001; Lawson *et al.*, 2004) (Fig. 19 B). La molécule d'ADN contient alors deux zones de courbures, l'une entre les nucléotides 6 et 7 (primaire) et l'autre entre les nucléotides -1 et 1 (secondaire) (Parkinson *et al.*, 1996). Les zones de courbure primaire sont dues en majorité à la compression du sillon majeur de l'ADN, alors que les zones de courbure secondaire sont plutôt dues à une compression du sillon mineur de l'ADN. La courbure totale est due majoritairement à ces zones de courbure primaire. La séquence au niveau des courbures (positions 6 et 7) n'est pas essentielle pour que la courbure soit mise en place lors de l'interaction avec le complexe CRP-AMPc, ce qui implique que c'est l'interaction protéine-ADN qui permet de générer la courbure (Chen *et al.*, 2001b; Chen *et al.*, 2001a). Néanmoins, l'angle de la courbure induite par la fixation du complexe varie en fonction de la séquence de l'ADN (Gartenberg et Crothers, 1988; Meyer-Almes et Porschke, 1997; Kahn et Crothers, 1998). En effet, Pyles et Lee (1998b) ont fait varier la séquence du site de fixation de CRP-AMPc et ont étudié la courbure générée. Les auteurs ont ainsi montré que la courbure pouvait être asymétrique et que l'angle de courbure dépendait des séquences flanquantes (Pyles et Lee, 1998b).

L'efficacité de fixation du complexe CRP-AMPc à l'ADN dépend donc de la séquence pentamérique inversée répétée, mais aussi des séquences flanquantes et des séquences séparant les séquences palindromiques.

II.3.2.1.3) Modulation de la capacité de fixation par phosphorylation.

Il existe une famille de facteurs de la transcription dont l'activité est régulée par formation d'une liaison covalente notamment par phosphorylation. Il s'agit des systèmes à deux composants ou phospho-relais. Ils sont constitués par deux partenaires (pour revue : Egger *et al.*, 1997; Hoch, 2000) : une protéine histidine kinase membranaire (HK) et une protéine qui joue le rôle de régulateur de réponse (RR). L'HK dont l'activité kinase est sensible à certaines conditions extérieures peut s'autophosphoryler et transmettre son phosphate au régulateur RR. Cette protéine ainsi phosphorylée acquiert la capacité de se fixer à l'ADN et de réguler la transcription. Un exemple est le système phospho-relais EnvZ (HK) et OmpR (RR) (Egger *et al.*, 1997).

II.3.2.1.4) Modulation de la capacité de fixation par séquestration.

Lors du transport du glucose à l'intérieur de la cellule par le système PTS, le répresseur de transcription Mlc est séquestré au niveau de la membrane cytoplasmique (pour revue : Plumbridge, 2002). De ce fait, la concentration intracellulaire en facteur actif libre diminue et n'est pas suffisante pour réprimer la transcription des gènes, en particulier, ceux qui permettent l'expression des protéines du système PTS.

Un autre exemple de séquestration de facteur de transcription au niveau de la membrane interne est celui du facteur de transcription MalT, qui régule l'expression des gènes *malEFG* et *malk*, qui permettent la synthèse du transporteur ABC du maltose et des maltodextrines. En absence de ces substrats dans le milieu de culture, la protéine MalK qui est située au niveau de la membrane permet de fixer l'activateur transcriptionnel MalT. La transcription des gènes *malEFG* et *malk* s'effectue donc à un niveau basal. En présence de maltose, le transporteur est activé (le maltose est incorporé dans la cellule) et MalT ne se fixe plus au niveau de MalK. Il permet ainsi une activation de l'initiation de la transcription au niveau de ces gènes (pour revue : Bohm et Boos, 2004).

II.3.2.2) Régulation négative.

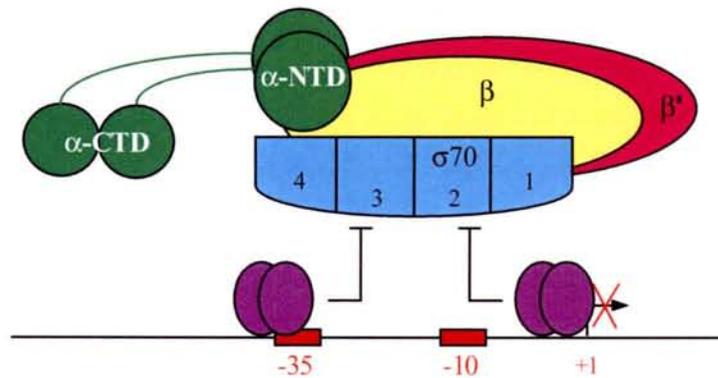
Les capacités de fixation à l'ADN d'un répresseur sont régulées par les différents paramètres décrits précédemment et dans le cas d'une régulation négative, cette fixation à l'ADN va limiter l'efficacité d'initiation de la transcription par l'ARN pol. Le taux d'occupation des opérateurs par ces répresseurs est un paramètre très important de la régulation. En effet, plus ce taux est élevé, plus la répression est efficace. Ces répresseurs agissant au niveau de l'initiation de la transcription ont en général un site de fixation à l'ADN qui chevauche le site de fixation de l'ARN pol, ce qui bloque soit la liaison de l'holoenzyme au promoteur, soit la formation du complexe RPo (Rojo, 2001; Rojo, 1999; Madan Babu et Teichmann, 2003a). Certains répresseurs limitent l'efficacité d'élongation en se fixant en aval du promoteur et en bloquant la progression de l'holoenzyme. Il existe aussi des répresseurs dont les sites de fixation ne chevauchent pas le site de fixation de l'ARN pol, mais qui affectent d'autres étapes du mécanisme d'initiation de la transcription, en se fixant à proximité des régions promotrices (pour revue : Choy et Adhya, 1996). Les interactions qui peuvent être mises en place entre les répresseurs et les sous-unités de l'ARN pol sont répertoriées dans la figure 18. Les répresseurs de classe I entrent en interaction avec le domaine α -CTD de l'ARN pol.

Il est à noter que les facteurs antisigma décrits plus haut peuvent aussi être classés dans les répresseurs transcriptionnels, puisqu'ils inhibent la première étape de l'initiation de la transcription qui est l'assemblage de l'holoenzyme (Rojo, 1999). Ces facteurs antisigma peuvent aussi inhiber l'activité des facteurs σ en masquant les sites de reconnaissance de l'ADN (pour revue : Hughes et Mathee, 1998).

Plusieurs mécanismes ont été décrits pour les différents types de répression de la transcription (Fig. 20) (Choy et Adhya, 1996; Rojo, 2001; Rojo, 1999; Browning et Busby, 2004). Dans le cas le plus simple, le site de fixation du répresseur chevauche les sites de reconnaissance de l'ARN pol, l'holoenzyme ne peut pas se fixer et initier la transcription (Fig. 20 A). Plusieurs exemples correspondant à ce mécanisme ont été décrits, notamment l'action du répresseur LacI par fixation au site O1 du promoteur *lac*. Ce site de fixation chevauche le site d'initiation de la transcription (centré en +11 par rapport au site +1) (Schlax *et al.*, 1995; Muller-Hill, 1998).

Un autre exemple de protéine répondant à ce mécanisme est celui de la protéine Mlc qui se fixe au niveau des promoteurs contrôlant l'expression des enzymes du système PTS. Ainsi, le

A



B

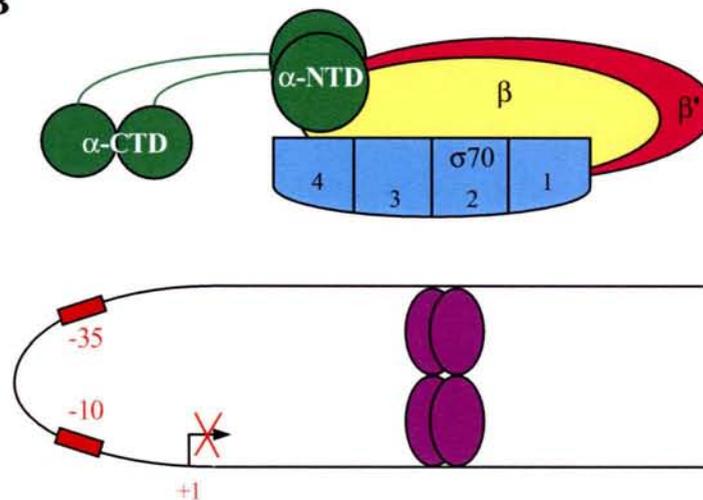


Figure 20 : Les différents mécanismes de répression de l'initiation de la transcription (d'après Brownig et Busby, 2004).

A) Répression par encombrement stérique. Les sites de fixation du répresseur (violet) se situent au niveau des sites de reconnaissance de l'ARN pol.

B) Répression par formation d'une boucle d'ADN. Les répresseurs se fixent sur des séquences distales et leur interaction permet la formation d'une boucle d'ADN qui empêche la fixation de l'ARN pol.

site de fixation de Mlc chevauche la séquence de la boîte -10 et le site d'initiation de la transcription du promoteur du gène *ptsG* codant le transporteur spécifique du glucose EIIBC^{Glc} (Plumbridge, 1998b; Plumbridge, 2002).

Les protéines intervenant dans la structuration de l'ADN (Fis, H-NS, HU, IHF, StpA, Dps) peuvent aussi être des répresseurs de la transcription, lorsque leurs sites de fixation chevauchent les séquences d'un promoteur (McLeod et Johnson, 2001).

Certains répresseurs agissent en induisant la formation d'une boucle de l'ADN (Fig. 20 B). En effet, lorsque les sites de fixation des répresseurs se situent de part et d'autre du promoteur, l'interaction de deux molécules fixées sur leurs séquences cibles conduit à la formation d'une boucle de l'ADN qui bloque l'accès de l'ARN pol au promoteur. Plusieurs exemples de répressions ont été décrits, c'est notamment le cas pour le répresseur GalR qui se fixe sur les promoteurs du gène *gal* (Choy et Adhya, 1992; Choy et Adhya, 1996) et aussi sur le promoteur *lac* (Borowiec *et al.*, 1987). En effet, en plus de l'opérateur O1, il existe deux éléments LacI qui permettraient la formation d'une boucle de l'ADN. D'une manière générale, l'existence de courbures de l'ADN entre les deux sites de fixation des répresseurs et le taux de surenroulement peuvent influencer la formation de ces boucles de l'ADN. Le répresseur NagC qui limite l'efficacité des promoteurs divergents *nagE* et *nagBAC* et contrôle l'expression des gènes impliqués dans le transport et la dégradation du N-acetylglucosamine, permet aussi de créer des boucles dans l'ADN (Plumbridge, 2001; Choy et Adhya, 1996).

La répression ne passe pas forcément par le blocage de la fixation de l'ARN pol, on peut avoir fixation du répresseur sur son opérateur en même temps que la fixation de l'ARN pol sur le promoteur mais dans ce cas, l'effet inhibiteur implique le ralentissement d'une étape ultérieure du mécanisme d'initiation (Roy *et al.*, 1998; Rojo, 1999; Rojo, 2001). Ce type de régulation est utilisé par le répresseur IclR qui affecte l'expression des gènes *aceBAK* codant l'isocitrate lyase (*aceB*), la malate synthase (*aceA*) et l'isocitrate déshydrogénase kinase/phosphorylase (*aceK*) du shunt du glyoxylate (Yamamoto et Ishihama, 2003). Au niveau du promoteur *aceB*, deux sites de fixation pour IclR ont été détectés. Le premier site, centré à la position -38, conduit à une inhibition classique par compétition entre la fixation du facteur et de l'ARN pol (Negre *et al.*, 1992) mais il existe un deuxième site localisé entre les positions -125 et -99 par rapport au site +1. IclR fixé au niveau de ce site interagit avec la sous-unité α -CTD de l'ARN pol, ce qui déstabilise le complexe RPo et entraîne une inhibition de la transcription (Yamamoto et Ishihama, 2003). D'autres exemples impliquant la sous-unité α ont été décrits pour les répresseurs GalR (Choy *et al.*, 1995) et PhoP (Yamamoto *et al.*, 2002).

Dans d'autres cas, la fixation du répresseur pourrait limiter l'échappement de l'ARN pol. En effet, si l'étape limitante de l'initiation est l'échappement de l'ARN pol, un facteur de transcription qui stabilise les interactions entre l'ARN pol et le promoteur en interagissant avec le domaine α -CTD, peut renforcer la limitation à l'étape d'échappement (Rojo, 1999; Rojo, 2001).

II.3.2.3) Régulations positives.

Dans le cas d'une régulation positive, le promoteur qui dirige l'expression du gène à réguler est en général sous-optimale. L'activateur va améliorer l'étape limitante et aussi favoriser l'initiation de la transcription (Roy *et al.*, 1998).

L'influence positive que peut avoir un facteur protéique sur l'initiation de la transcription dépend de plusieurs paramètres : la position qu'il occupe sur l'ADN par rapport au promoteur et le taux d'occupation de son site de fixation, ainsi que les interactions qu'il entretient avec l'ARN pol. Deux grandes classes de facteurs protéiques ont ainsi été mises en évidence. Il s'agit des activateurs de classe I et de classe II (Fig. 18 et 21) (pour revue : Browning et Busby, 2004; Busby et Ebright, 1999; Busby et Ebright, 1994). Ces classes sont définies par la position du site de fixation de l'activateur par rapport au site d'initiation de la transcription.

II.3.2.3.1) Les activateurs de classe I.

Les sites de fixation des activateurs de classe I sont centrés aux positions -61,5, -72,5, -82,5 et -92,5 par rapport au site +1 d'initiation de la transcription, ce qui fait qu'ils sont tous situés sur la même face de la double hélice (Fig. 21 A et B) (Busby et Ebright, 1999; Lawson *et al.*, 2004; Browning et Busby, 2004).

L'exemple le plus connu de promoteur activé par un activateur de la classe I est le promoteur *lac* dont l'initiation de la transcription est activée par le complexe CRP-AMPc (Ebright, 1993). Le site de fixation de CRP-AMPc est centré à la position -61,5 (Gaston *et al.*, 1990).

Lorsque l'activateur est fixé sur son site, il facilite le recrutement de l'ARN pol en interagissant avec le domaine CTD de la sous-unité α . La région reliant les sous-unités α -CTD et α -NTD est flexible, ce qui permet le positionnement des activateurs aux positions -61,5, -72,5, -82,5 et -92,5. Le mode d'action du complexe CRP-AMPc serait de faciliter la fixation de l' α -CTD sur l'ADN, ce qui favoriserait le recrutement de l'ARN pol au niveau du

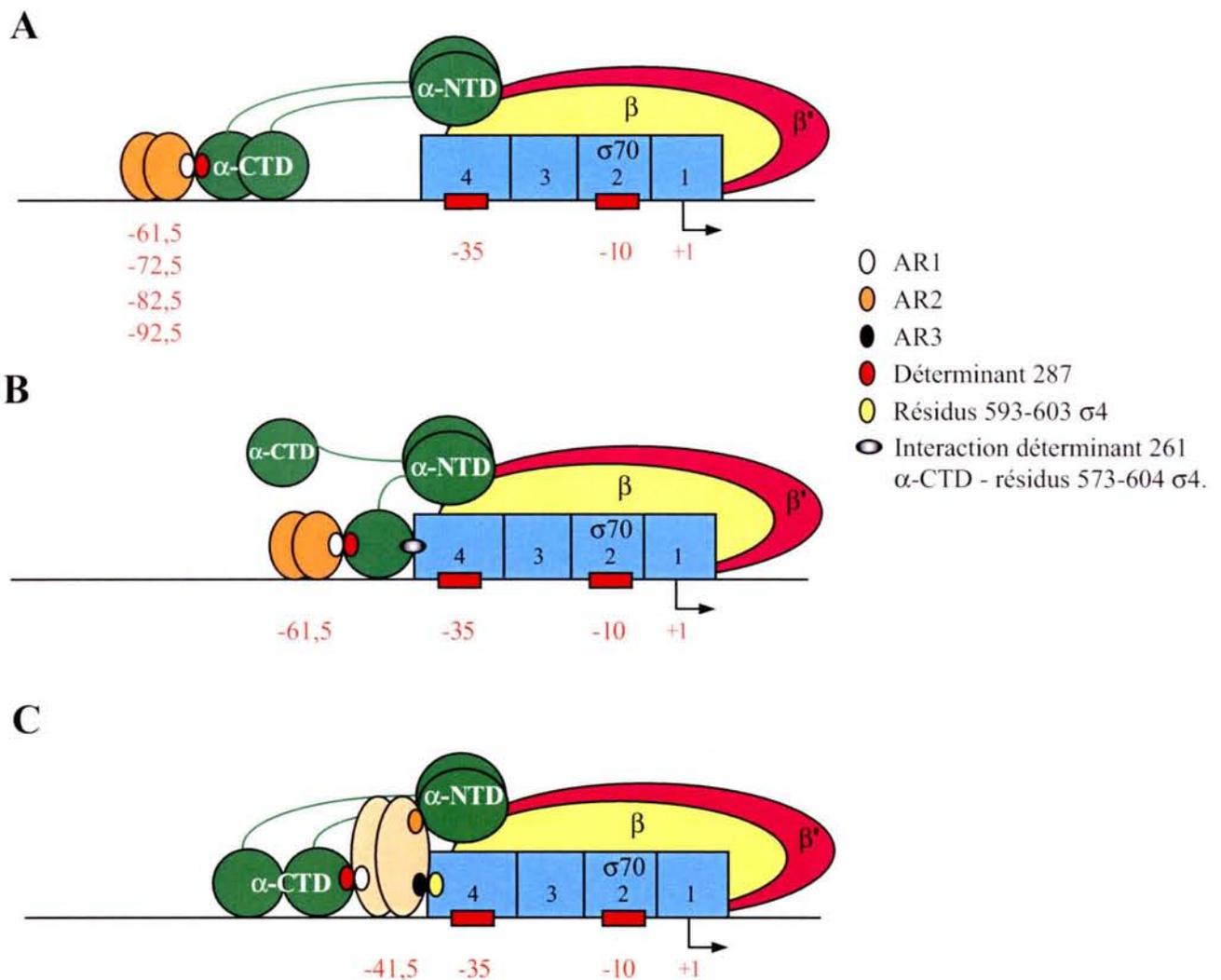


Figure 21 : Les différents mécanismes d'activation de l'initiation de la transcription (d'après Browning et Busby, 2004 ; Lawson *et al.*, 2004; Busby et Ebright, 1994). La plupart des activateurs agissent sous forme de dimères.

A) Activateurs de classe I (exemple du promoteur *lac* P1). Le monomère proximal de l'activateur interagit avec le déterminant 287 du domaine α -CTD de l'ARN pol par l'intermédiaire de la région AR1. La position possible des sites de fixation des activateurs sur l'ADN par rapport au site +1 est indiquée.

B) Dans le cas du complexe CRP-AMPc (classe I), une interaction entre le déterminant 261 de l' α -CTD et le domaine 4 de la sous-unité σ de l'ARN pol a été démontrée (Chen *et al.*, 2003). Cette interaction ne peut se produire que lorsque l'opérateur est situé à la position -61,5 par rapport au site d'initiation de la transcription.

C) Activateurs de classe II. Le site de fixation de l'activateur se situe en position -41,5 par rapport au site +1 d'initiation de la transcription. Dans le cas du complexe CRP-AMPc (classe II ; exemple du promoteur *gal* P1), des interactions entre la région AR2 et les résidus 162-165 de l' α -NTD et la région AR3 avec la région 593-603 du domaine σ 4 de l'ARN pol ont été mises en évidence.

promoteur (Lawson *et al.*, 2004). Toujours au niveau du promoteur *lac* P1, un autre rôle du complexe CRP-AMPc serait de faciliter l'étape d'isomérisation (Liu *et al.*, 2003). Cet effet serait plus faible que l'influence sur la liaison de l'ARN pol au promoteur.

La région de la protéine CRP qui est en interaction avec l' α -CTD a été mise en évidence grâce à l'étude de mutants de CRP qui peuvent se fixer à l'ADN mais qui ne peuvent plus activer la transcription. Cette région de fixation qui a été appelée AR1 (activating region 1) est localisée entre les résidus 156 et 164 de CRP (Fig. 21 A) (Bell *et al.*, 1990; Eschenlauer et Reznikoff, 1991; Niu *et al.*, 1994; Zhou *et al.*, 1993a) (pour revue : Busby et Ebright, 1999). Les contacts entre l'ARN pol et CRP pourraient se mettre en place par l'intermédiaire de la région AR1 de la sous-unité proximale du dimère de CRP (Zhou *et al.*, 1993b).

De la même manière, les acides aminés de l' α -CTD qui interagissent avec CRP ont été identifiés par la technique d'alanine « scanning » qui consiste à remplacer chaque acide aminé d'une région de la protéine en alanine. Les effets produits sur l'activation ont permis de montrer que les résidus de l' α -CTD appelés « déterminant 287 » (Tyr 285, Glu 286, Val 287, Glu 288 et Arg 317), « déterminant 265 » et « déterminant 261 » sont impliqués dans l'activation par le complexe CRP-AMPc (Savery *et al.*, 2002).

La région AR1 de CRP-AMPc interagirait avec le déterminant 287, alors que le déterminant 265 permettrait la fixation de l' α -CTD sur l'ADN (Fig. 21 A) (Savery *et al.*, 2002).

Une interaction supplémentaire a été détectée au niveau du promoteur *lac*. En effet, l' α -CTD dont la fixation sur l'ADN est favorisée par son interaction avec la région AR1 de CRP-AMPc peut interagir avec le domaine $\sigma 4$ (résidus 573-604) de l'ARN pol (Fig. 21 B) (Chen *et al.*, 2003). La région de l' α -CTD qui serait importante pour cette interaction serait le déterminant 261 (Chen *et al.*, 2003 ; Lawson *et al.*, 2004). Cette interaction supplémentaire ne serait possible que si le site de fixation de l'activateur est centré en -61,5 par rapport au site +1. Cette interaction pourrait faciliter une autre étape du mécanisme d'initiation de la transcription, notamment l'isomérisation de l'ARN pol pour obtenir le complexe RPo.

Une structure du complexe CRP-AMPc, ADN et α -CTD a été établie par le groupe de R. Ebright, ce qui a permis de confirmer l'identité des acides aminés impliqués dans les interactions décrites précédemment (Fig. 22 A) (Benoff *et al.*, 2002; Lawson *et al.*, 2004). Cette équipe a ensuite modélisé l'interaction entre le complexe CRP-AMPc, l' α -CTD et l'ARN pol fixés sur une molécule d'ADN (Lawson *et al.*, 2004). Cette modélisation montre l'organisation structurale potentielle d'un activateur de classe I par rapport à l'ARN pol, notamment le domaine α -CTD qui pourrait interagir à la fois avec CRP-AMPc et le domaine

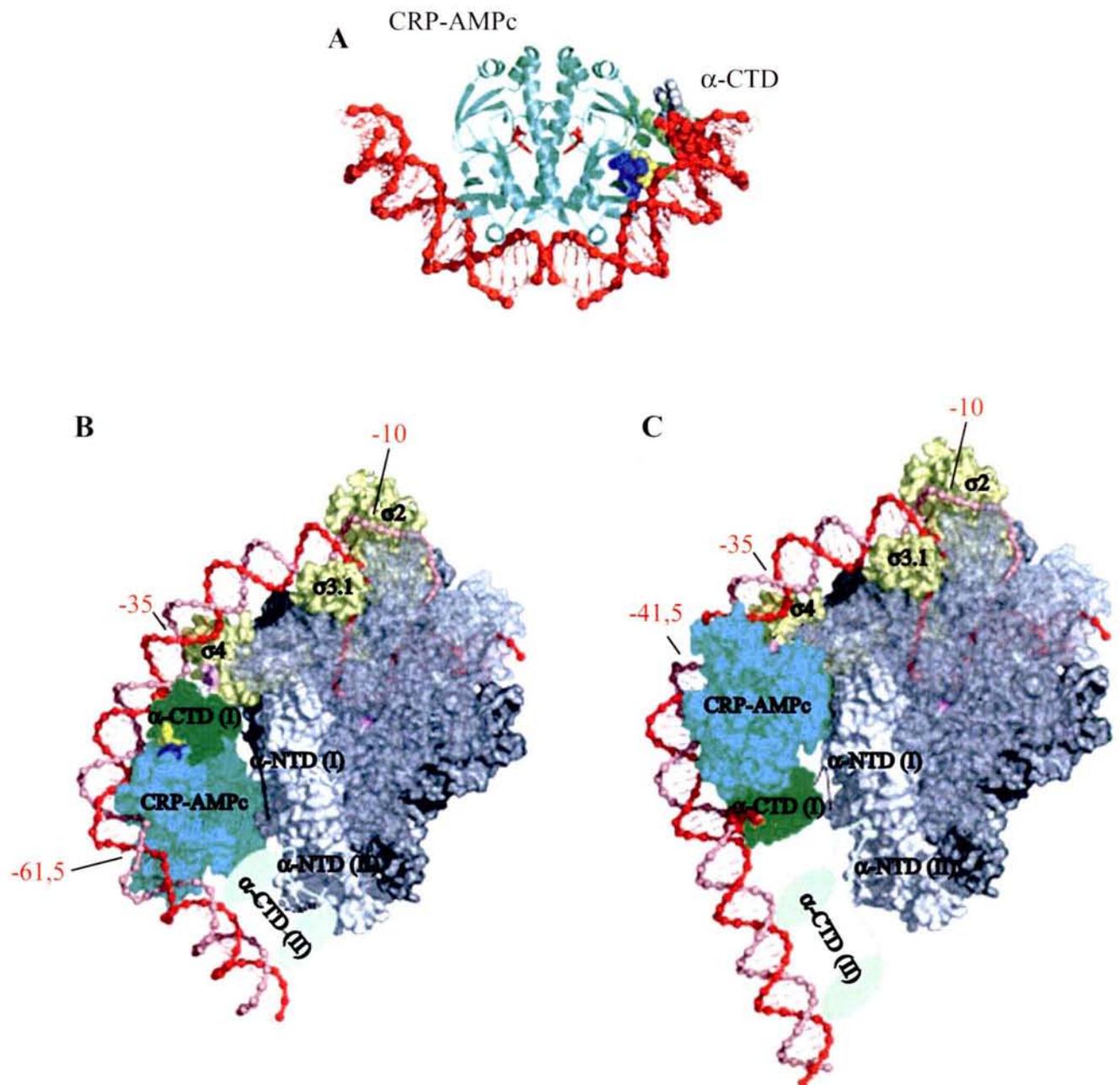


Figure 22 : Modèles d'activation par le complexe CRP-AMPc (d'après Lawson *et al.*, 2004).

A) Structure 3D du complexe CRP-AMPc en interaction avec le domaine α -CTD fixé et une molécule d'ADN (Benoff *et al.*, 2002). Les interactions présentes dans ce complexe sont représentatives des activateurs de classe I et II. Le complexe CRP-AMPc est représenté en bleu clair avec l'AMPc en rouge. L' α -CTD est représenté en vert alors que la molécule d'ADN est représentée en rouge. La région AR1 de CRP-AMPc est représentée en bleu alors que les trois régions du domaine α -CTD qui peuvent être contactées par les activateurs sont représentées en rouge pour le déterminant 265, en blanc pour le déterminant 261 et en jaune pour le déterminant 287.

B et C) Modélisation des complexes CRP-AMPc, α -CTD et ARN pol fixés sur une molécule d'ADN dans le cas d'un activateur de classe I (B) et de classe II (C). L'ARN pol est représentée en gris (les intensités de gris changent suivant la sous-unité), le complexe CRP-AMPc est représenté en bleu et les domaines α -CTD (I) et α -CTD (II) en vert foncé et clair, respectivement. Les régions AR1, AR2 et AR3 de CRP-AMPc sont représentées en bleu foncé, vert foncé et vert clair et les déterminants de l' α -CTD sont représentés avec les mêmes couleurs qu'en A). Le brin matrice est représenté en rouge et le brin non matrice en rose. Les domaines du facteur σ sont indiqués en jaune. Le domaine α -CTD (II) est représenté à deux positions potentielles différentes sur l'ADN. La position des interactions est indiquée.

$\sigma 4$ (Fig. 22 B). Cette modélisation montre aussi l'enroulement de la molécule d'ADN autour de l'ARN pol.

Les données obtenues pour le complexe CRP-AMPc peuvent s'appliquer aux promoteurs qui possèdent un site de fixation pour un activateur de la transcription centré aux positions -61,5, -72,5, -82,5 et -92,5. Ainsi, la protéine Fis qui permet d'activer les promoteurs *rrn* P1 des opérons ribosomiques et d'autres promoteurs permettant la transcription d'ARN stables se fixe sur un site localisé à la position -71 par rapport au site +1, ce qui permet à Fis d'interagir avec l' α -CTD et aussi d'activer la transcription au promoteur *rrnB* P1. Pour le promoteur *tyrT*, Fis pourrait à la fois faciliter le déroulement de l'ADN au niveau de la boîte -10 et l'échappement de l'ARN pol lors du mécanisme d'initiation de la transcription en plus du recrutement de l'ARN pol au niveau du promoteur (Muskhelishvili *et al.*, 1997; Auner *et al.*, 2003).

II.3.2.3.2) Les activateurs de classe II.

En général, les sites de fixation des activateurs de classe II chevauchent la séquence de la boîte -35. En effet, pour ce type de promoteur, la séquence de la boîte -35 est divergente par rapport à la séquence consensus (pour revue : Busby et Ebright, 1997; Busby et Ebright, 1994; Busby et Ebright, 1999; Lawson *et al.*, 2004; Browning et Busby, 2004). Ces sites sont centrés à la position -41,5 par rapport au site d'initiation de la transcription (Fig. 21 C) et permettent une interaction des facteurs avec la sous-unité σ de l'ARN pol. Cette interaction pourrait augmenter l'affinité de l'ARN pol pour le promoteur mais aussi influencer d'autres étapes du mécanisme d'initiation de la transcription. Du fait de l'interaction entre l'activateur et le facteur σ , les sites de fixation des activateurs de classe II sont toujours situés à proximité de la boîte -35.

Le rôle des activateurs de classe II a été étudié en détail sur deux promoteurs : le promoteur *galP1* activé par le complexe CRP-AMPc et le promoteur *CC(-41,5)* activé par le complexe CRP-AMPc (pour revue : Busby et Ebright, 1997; Busby et Ebright, 1999) (Fig. 21 C).

Pour ces deux promoteurs, le complexe CRP-AMPc contacte l'ARN pol en de très nombreux points (Fig. 21 C) (pour revue : Busby et Ebright, 1997; Busby et Ebright, 1999; Lawson *et al.*, 2004). Le site de fixation du complexe CRP-AMPc est positionné en -41,5 par rapport au site +1 et chevauche la séquence de la boîte -35. L'activateur CRP-AMPc contacte l' α -CTD (déterminant 287) avec la région AR1 comme dans le cas des activateurs de classe I (Bell *et*

al., 1990; Zhou *et al.*, 1994a) mais au niveau de sa sous-unité distale (Zhou *et al.*, 1994b). La sous-unité proximale de CRP-AMPc interagit aussi avec la sous-unité α -NTD (résidus 162-165) par l'intermédiaire d'une autre région AR2 (activating region 2) (Niu *et al.*, 1996). En plus des régions AR1 et AR2, il peut exister un troisième type d'interaction entre la région AR3 de CRP-AMPc et le domaine σ 4 (résidus 593-603) de l'ARN pol (Niu *et al.*, 1996). Cette interaction ne se produit pas en présence d'une région AR3 sauvage pour CRP-AMPc (Niu *et al.*, 1996) mais est essentielle pour un homologue de CRP : la protéine Fnr (Lonetto *et al.*, 1998).

Comme dans le cas des activateurs de classe I, l'interaction de AR1 avec l' α -CTD permet d'augmenter l'affinité de l'ARN pol pour le promoteur (Rhodius *et al.*, 1997). Par contre, le contact entre AR2 et l' α -NTD pourrait favoriser la formation des complexes RPo (Rhodius *et al.*, 1997) (pour revue : Busby et Ebright, 1997; Busby et Ebright, 1999; Lawson *et al.*, 2004). Le contact AR3/ σ 4 pourrait aussi favoriser la formation des complexes ouverts (Busby et Ebright, 1999; Rhodius et Busby, 2000). Les interactions entre le complexe CRP-AMPc et l'ARN pol fixés sur une molécule d'ADN ont été modélisés dans le cas d'un activateur de classe II (Fig. 22 C) (Lawson *et al.*, 2004). Cette modélisation confirme les positions des zones d'interaction entre les différents partenaires.

II.3.2.4) Modèles complexes.

D'autres modèles plus complexes ont été établis pour expliquer les régulations observées au niveau de certains promoteurs. Dans ce cas, la régulation est très dépendante de la concentration du facteur de transcription du fait de la présence de différents sites de fixation ayant des affinités différentes pour ce facteur, ceci en amont du promoteur. L'occupation de ces différents sites permet la régulation de l'initiation de la transcription. Ce type de régulation fait en général intervenir des protéines qui peuvent modifier la structure de l'ADN, notamment la courbure de l'ADN, ce qui permet une modulation de la fixation des facteurs de transcription et/ou de l'ARN pol en fonction des conditions de croissance et ainsi une régulation de l'initiation de la transcription en fonction de ces conditions. Les régulations complexes de ce type les plus observées chez *E. coli* font souvent intervenir la fixation de plusieurs régulateurs, ce qui permet l'intégration de plusieurs signaux.

II.3.2.4.1) Régulation reposant sur un taux d'occupation différentiel des séquences régulatrices.

L'expression du gène *crp* codant la protéine CRP est autorégulée grâce à l'existence de deux sites de fixation pour le complexe CRP-AMPc, présents au niveau de la région promotrice et ayant une affinité différente pour le complexe CRP-AMPc (Hanamura et Aiba, 1992; Aiba, 1983). Le premier site (I) est centré à la position -61,5 par rapport au site d'initiation du promoteur P1, tandis que le deuxième site (II) est situé en aval du promoteur à la position +42. Il est à noter que le site II a une affinité plus forte pour CRP-AMPc que le site I (Pyles et Lee, 1996). En présence d'une concentration faible en complexe CRP-AMPc, le site II est majoritairement occupé, ce qui conduit à une inhibition de l'initiation de la transcription, alors que pour une concentration plus élevée en complexe, une activation de l'initiation à partir du promoteur *crp* P1 est possible grâce au complexe CRP-AMPc fixé au site I, qui jouerait un rôle d'activateur de classe I et recruterait l'ARN pol par interaction avec le domaine α -CTD.

II.3.2.4.2) Régulation par modification de la structure de l'ADN.

Comme nous l'avons vu, la structure de l'ADN a une forte influence sur l'efficacité de l'initiation de la transcription et les protéines Fis, IHF, H-NS et HU qualifiées de protéines histone-like contribuent à la structuration de l'ADN génomique bactérien (Azam et Ishihama, 1999) (pour revue : Pettijohn, 1996; Dorman et Deighan, 2003; Dorman, 2004). Un grand nombre de régulateurs transcriptionnels, tel que le complexe CRP-AMPc, modifient aussi la structure de l'ADN lors de leur fixation.

En changeant la conformation de l'ADN, ces protéines peuvent influencer la transcription soit directement dans le cas où le changement de conformation a un effet direct sur la formation du complexe d'initiation, soit indirectement en modulant l'efficacité de fixation d'un activateur ou d'un répresseur transcriptionnel (Fig. 23) (pour revue : Dai et Rothman-Denes, 1999; Perez-Martin et de Lorenzo, 1997; McLeod et Johnson, 2001).

Par exemple, au niveau du promoteur *narG*, l'induction d'une courbure par la protéine IHF favorise l'interaction d'un deuxième activateur NarL avec l'ARN pol (Fig. 23 A). Le positionnement optimal de NarL en présence d'IHF permet une activation de la transcription efficace (Schroder *et al.*, 1993).

Les promoteurs divergents *PmalE* et *PmalK* sont activés par la présence de l'activateur MalT et du complexe CRP-AMPc (pour revue : Richet, 1996; Perez-Martin et de Lorenzo, 1997).

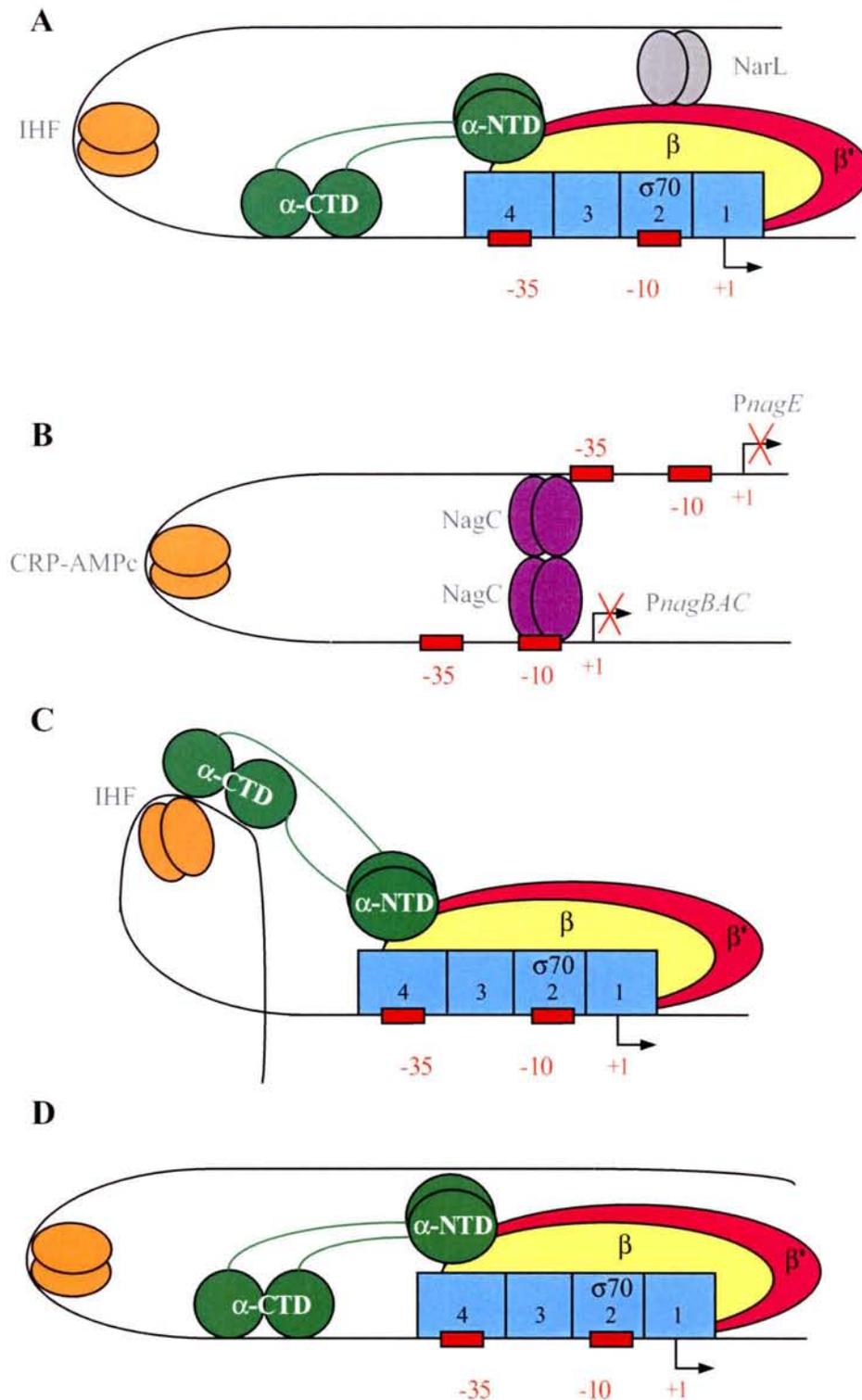


Figure 23 : Les protéines qui modifient la structure de l'ADN influencent l'initiation de la transcription (d'après Perez-Martin *et al.*, 1994 ; Perez-Martin et De Lorenzo, 1997).

A) L'induction d'une courbure par un facteur protéique peut renforcer l'interaction entre un activateur et l'ARN pol. Ainsi, la courbure induite par IHF facilite l'interaction de NarL avec $E\sigma^{70}$ au niveau du promoteur du gène *narG* (Schroder *et al.*, 1993).

B) L'induction d'une courbure par la protéine CRP-AMPc permet de favoriser l'interaction entre les deux répresseurs NagC au niveau des promoteurs divergents *nagE* et *nagBAC* (Plumbridge et Kolb, 1998).

C) La protéine IHF induit la formation d'une courbure de l'ADN en amont du promoteur λ PL qui permet de rapprocher la séquence d'un élément UP de l'ARN pol. La sous-unité α -CTD peut donc se fixer sur cette séquence et activer l'initiation de la transcription (Giladi *et al.*, 1998).

D) La fixation d'un facteur protéique qui induit la formation d'une courbure en amont d'un promoteur facilite l'enroulement de la molécule d'ADN autour de l'ARN pol.

Les sites de fixation de CRP-AMPc sont situés aux positions -76,5, -105,5 et -139,5 par rapport au site d'initiation du promoteur *PmalE*. Le complexe fixé au niveau de ces sites permet l'induction d'une courbure de l'ADN, ce qui entraîne un repositionnement des activateurs MalT au niveau d'autres sites de plus faible affinité. A ces nouvelles positions, MalT active l'initiation de la transcription aux deux promoteurs (Richet, 2000; Richet, 1996). Il est à noter que CRP-AMPc fixé au niveau du site centré à la position -139,5 pourrait aussi interagir directement avec l' α -CTD de l'ARN pol pour activer l'initiation de la transcription du promoteur *PmalE* (Richet, 2000). De plus, dans le cas des promoteurs divergents *PnagBAC* et *PnagE*, la fixation du complexe CRP-AMPc permet la formation d'une courbure de l'ADN qui favorise l'interaction entre deux molécules de répresseurs NagC liées à deux sites distincts (Fig. 23 B) (Plumbridge et Kolb, 1993; Plumbridge et Kolb, 1998).

Dans le cas du promoteur λP_L , la protéine IHF qui courbe fortement l'ADN lors de sa fixation permet de rapprocher la séquence d'un élément UP qui se lie à l' α -CTD de l'ARN pol (Fig. 23 C). Cette interaction permettrait d'augmenter l'affinité de l'ARN pol pour ce promoteur (Giladi *et al.*, 1998).

La protéine Fis possède trois sites de fixation en amont du promoteur *leuV* qui permet d'initier la transcription de l'opéron *leuV* codant les ARNt Leu (pour revue : Hatfield et Benham, 2002). A la position -71, la protéine Fis interagit avec l' α -CTD de l'ARN pol, alors que sa fixation au niveau des sites en amont permettrait la « transmission » d'une distorsion de l'ADN due au surenroulement vers la région qui est ouverte lors de la formation du complexe RPo (Opel *et al.*, 2004).

La fixation d'une protéine qui permet la courbure de l'ADN peut aussi favoriser l'enroulement de la double hélice autour de l'ARN pol favorisant ainsi la formation du complexe RPo (Fig. 23 D) (Perez-Martin *et al.*, 1994).

II.3.2.4.3) Intervention de plusieurs régulateurs et intégration des signaux.

L'expression d'un grand nombre de gènes est modulée par un seul régulateur transcriptionnel mais l'expression de 49% des gènes serait régulée par plusieurs régulateurs, dont les régulateurs globaux (CRP-AMPc, ArcA, Fis, IHF, FNR, NarL et Lrp) et des régulateurs plus spécifiques (Martinez-Antonio et Collado-Vides, 2003). La régulation par au moins deux régulateurs permet d'intégrer plusieurs signaux différents qui aboutissent à une initiation de la

transcription optimale en fonction des conditions de croissance de la bactérie. Par ailleurs, plus le nombre de régulateurs est élevé, plus le nombre de signaux intégrés est élevé et plus la régulation est fine (pour revue : Barnard *et al.*, 2004; Browning et Busby, 2004).

Il existe peu d'exemples de promoteurs régulés uniquement par des répresseurs, alors qu'un grand nombre de régulations impliquent la présence d'un activateur et d'un répresseur ou de deux activateurs (Fig. 24).

Lors de la régulation de la transcription par deux activateurs, deux mécanismes sont possibles. En effet, un activateur de classe I peut augmenter l'affinité de l'ARN pol pour le promoteur en interagissant avec l' α -CTD, alors qu'un autre activateur peut interagir avec la sous-unité σ (classe II). Le promoteur *proP2* est un exemple de ce type, où Fis fonctionne comme un activateur de classe II (positionné en -41,5) et CRP-AMPc positionné en amont (-121,5) fonctionne comme un activateur de classe I (McLeod *et al.*, 2002). Il est à noter que ce promoteur est reconnu par l'holoenzyme $E\sigma^S$. Belyaeva *et al.* (1998) ont utilisé un promoteur hybride reconnu par $E\sigma^{70}$ qui contient deux sites de fixation pour CRP-AMPc. Le premier est positionné en -41,5 (classe II) alors que le deuxième est positionné à différents endroits. Les auteurs ont montré que CRP-AMPc fixé aux sites en amont activait l'initiation de la transcription en interagissant avec l' α -CTD (classe I) (Belyaeva *et al.*, 1998).

Le deuxième mécanisme possible implique deux activateurs qui recrutent séparément les deux sous-unités α -CTD (Fig. 24 B). Les activateurs sont donc tous les deux de la classe I. Le promoteur *acsP2* dont l'initiation de la transcription est activée par deux molécules de CRP-AMPc positionnées en amont du promoteur (positions -69,5 et -122,5) est un exemple de ce type de régulation (Beatty *et al.*, 2003).

Dans le cas de la fixation d'un activateur et d'un répresseur au niveau d'un même promoteur, plusieurs mécanismes sont possibles. Si les sites de fixation des deux protéines se chevauchent, c'est la concentration et/ou l'activité de ces régulateurs qui va permettre soit l'activation, soit la répression de l'initiation de la transcription. Par contre, si les sites de fixation des régulateurs ne se chevauchent pas, d'autres mécanismes sont possibles. Un des cas les plus étudiés est le promoteur du gène *lac*, dont l'initiation de la transcription est : 1) activée par la présence de CRP-AMPc dont la concentration varie en fonction de la présence de glucose dans le milieu de culture et 2) réprimée par le répresseur de l'opéron *lac*, LacI (Fig. 24 C). En présence de l'inducteur physiologique allolactose ou non physiologique, l'IPTG, la structure de LacI est modifiée et le répresseur ne se fixe plus à l'ADN. Le résultat

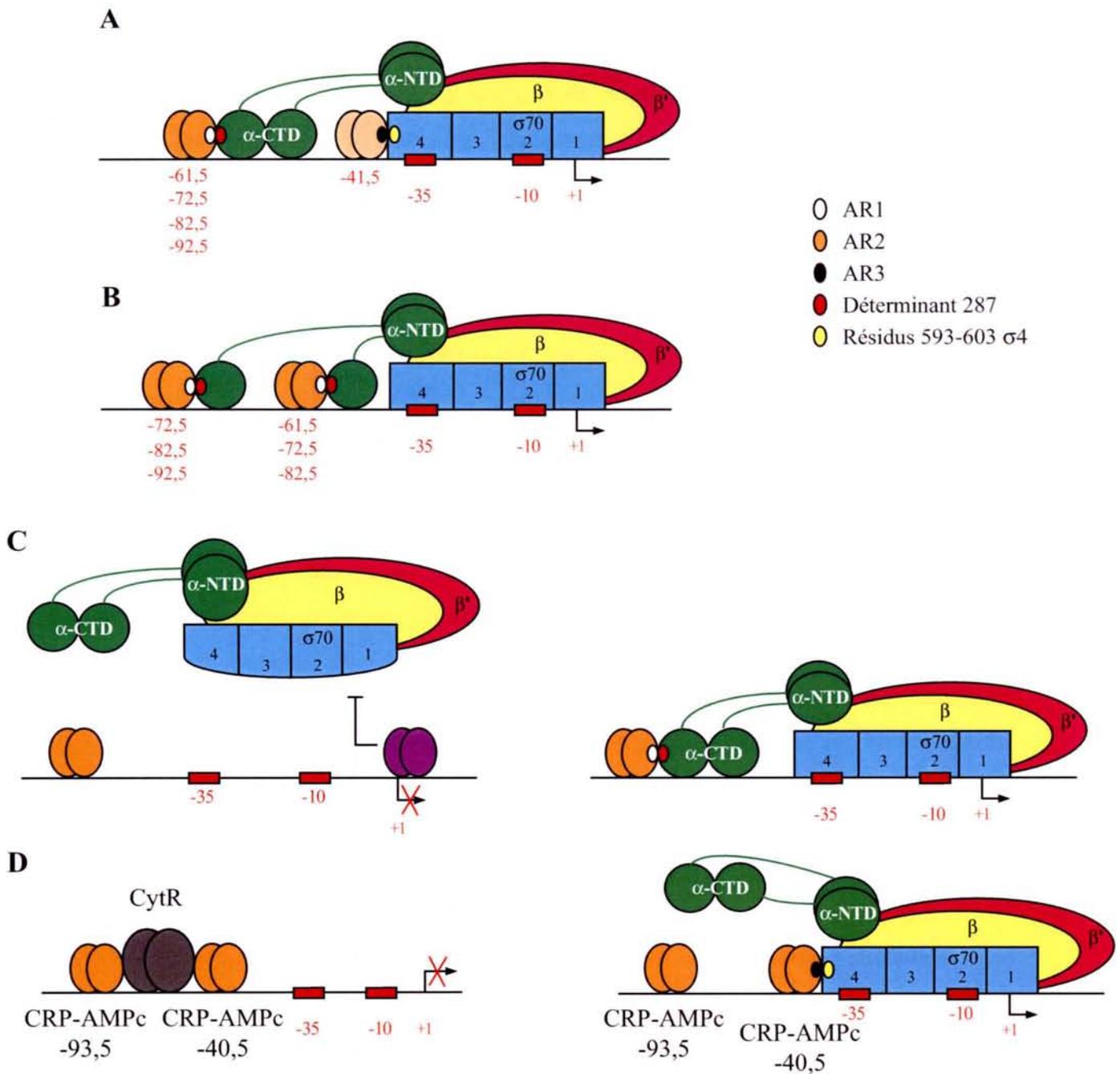


Figure 24 : Action simultanée de plusieurs régulateurs de la transcription (d'après Busby et Ebright, 1999, Lawson *et al.*, 2004 ; Barnard *et al.*, 2004 ; Browning *et al.*, 2004).

A) Contacts indépendants entre un activateur de classe I, un activateur de classe II et l'ARN pol. Le domaine α -CTD de l'ARN pol peut interagir avec un activateur fixé en amont (classe I) et un autre activateur fixé en position -41,5 peut interagir avec le domaine $\sigma 4$ de l'ARN pol. Les positions des sites de fixation des deux activateurs ainsi que les zones d'interaction entre les activateurs et l'ARN pol sont indiqués.

B) Deux activateurs distincts de classe I fixent les deux domaines α -CTD de façon indépendante. Les sites de fixation de deux activateurs ainsi que les zones d'interaction entre les activateurs et l'ARN pol sont indiqués.

C et D) Intégration de deux signaux différents pour optimiser l'initiation de la transcription (voir détails dans le texte). C) Exemple du promoteur *lac* P1. La fixation du répresseur empêche la fixation de l'ARN pol sur le promoteur. En absence du répresseur, l'initiation de la transcription est activée par le complexe CRP-AMPc. D) La fixation de la protéine CytR entraîne un changement de conformation du complexe CRP-AMPc incompatible avec une activation de la transcription.

est qu'en présence de lactose et en absence de glucose dans le milieu de culture, la transcription du gène *lac* permet la production de β -galactosidase pour consommer le lactose. La protéine CytR qui contrôle l'expression des gènes impliqués dans le métabolisme des nucléosides peut se fixer au niveau de promoteurs régulés par CRP-AMPc, par exemple au niveau du promoteur *deoP2* (pour revue Valentin-Hansen *et al.*, 1996). L'interaction directe entre CytR et deux molécules de CRP-AMPc fixées respectivement sur leur opérateur empêche l'activation de la transcription par CRP-AMPc (Fig. 24 D) (Shin *et al.*, 2001). Cette répression serait liée à un changement allostérique de CRP-AMPc, lors de l'interaction avec CytR, qui de ce fait ne pourrait plus interagir de façon optimale avec l'ARN pol (Shin *et al.*, 2001; Chahla *et al.*, 2003).

Pour d'autres promoteurs fortement régulés, plusieurs régulateurs dont l'activité est très dépendante de leur concentration mais aussi de leur niveau de phosphorylation ou de la liaison d'un cofacteur peuvent se fixer sur la séquence ADN en amont du promoteur pour optimiser l'initiation de la transcription en fonction des conditions environnementales. Le promoteur *nir* est un exemple de ce type de régulation complexe. Il est régulé par Fnr (dont l'activité est régulée par la présence d'oxygène), par Fis (dont la concentration est régulée par la phase de croissance), par NarL/P (dont l'activité est régulée par la présence d'ions nitrite/nitrate), et par les histone-like H-NS et IHF (pour revue : McLeod et Johnson, 2001; Barnard *et al.*, 2004).

Ces régulations peuvent être encore plus complexes dans le cas où la structure de l'ADN influence à la fois la fixation des différents régulateurs mais aussi de l'ARN pol. De plus, la présence de plusieurs promoteurs reconnus par des holoenzymes associées à d'autres facteurs σ , dont la fixation au promoteur est influencée de façon différentielle par plusieurs facteurs protéiques, permet de contrôler l'initiation de la transcription et donc la production de protéines en fonction des conditions environnementales. En effet, la majorité des régulations décrites précédemment peuvent être observées également pour d'autres facteurs σ .

Partie 2 : Adaptation d'*E. coli* aux changements environnementaux.

Les micro-organismes peuvent utiliser une grande variété de sources de carbone et peuvent s'adapter au milieu extérieur grâce à leur capacité d'assimilation de nutriments présents à de très faibles concentrations dans le milieu de croissance. Un grand nombre de systèmes de transport existent chez les bactéries pour utiliser de manière optimale toute trace de nutriment. *E. coli* est une bactérie à gram négatif, qui possède deux membranes : une membrane externe et une membrane interne. L'espace créé entre ces deux membranes constitue le périplasme.

La différence majeure entre ces deux membranes est que la membrane externe est plus perméable que la membrane interne du fait de la présence de nombreuses protéines appelées porines qui forment des pores dans la membrane. Elles permettent en général la diffusion passive de solutés et de petites molécules au travers de la membrane externe vers le périplasme. Lors d'un changement environnemental, la composition en porine de la membrane externe est modifiée, ce qui permet une adaptation des capacités de perméabilité. Dans un premier temps, je vais décrire les propriétés et les fonctions des porines d'*E. coli*.

Un rôle essentiel des protéines qui sont enchâssées dans la membrane interne est de transmettre les signaux qui vont permettre l'adaptation de la bactérie aux conditions extérieures. Les signaux émis par ces systèmes entraînent une réponse qui permet, par exemple, un changement du taux de synthèse des protéines, une régulation de l'activité de certaines enzymes mais aussi une modification du comportement de la bactérie comme sa motilité. La transmission du signal repose sur des mécanismes de phosphorylation des protéines. Ces systèmes ont été appelés phospho-relais. Un exemple de phospho-relais sera décrit. Il s'agit du système PTS d'*E. coli* (Phosphoénolpyruvate phosphoTransferase System), qui permet à la fois de détecter la présence d'un sucre et de le transporter au travers de la membrane interne de la cellule. Je détaillerai les composants des systèmes de transport du glucose, ainsi que les régulations au niveau métabolique et génétique qui découlent de son transport par le système PTS spécifique du glucose.

Enfin, je décrirai un autre exemple de système basé sur la phosphorylation des protéines. Il s'agit des systèmes à deux composants impliqués dans la surveillance de l'environnement de la bactérie. Ces systèmes permettent la mise en place d'une réponse qui permet l'adaptation aux nouvelles conditions extérieures.

I.) Les porines bactériennes.

Différentes familles de porines sont présentes au niveau de la membrane externe d'*E. coli*. Certaines porines permettent la diffusion facilitée de composés hydrophiles et montrent des spécificités pour les cations ou les anions. Par contre, la taille des composés qui passent au travers du pore ne doit pas dépasser certaines dimensions. Ces porines ont été appelées majeures (pour revue : Schulz, 1996; Nikaido, 1996; Schirmer, 1998; Koebnik *et al.*, 2000).

L'autre type de porine a une spécificité pour le transport de certains type de molécules. Elles ont été appelées porines spécifiques (pour revue : Schulz, 1996; Nikaido, 1996; Schirmer, 1998; Koebnik *et al.*, 2000).

I.1) Les porines majeures.

Cette famille est représentée par les porines OmpF, OmpC et PhoE chez *E. coli* (Nikaido, 1996). Ces porines forment des homotrimères qui traversent la membrane externe (Fig. 25). Chaque monomère forme un tonneau de 16 ou 18 brins β antiparallèles qui permettent la formation d'un pore (Fig. 25 A). Les feuillettes sont connectés du côté périplasma par des boucles courtes, alors que du côté extracellulaire, ces boucles sont beaucoup plus longues et de séquences variables. Cette variabilité rendrait plus difficile la reconnaissance de ces macromolécules par des antibiotiques ou des phages. Les porines forment une structure à trois pores (Fig. 25 B). La région hydrophobe qui se situe entre les monomères serait impliquée dans l'assemblage de la porine. La nature hydrophile des acides aminés qui composent l'intérieur du pore empêche l'entrée de composés hydrophobes mais facilite l'entrée des ions (Nikaido, 1996).

Une des propriétés commune à un grand nombre de porines est la présence d'une boucle interne qui bloque l'entrée du pore (Schirmer, 1998). Cette boucle remplit en partie l'entrée du pore, ce qui limite la taille des composés qui peuvent franchir la membrane externe (la limite supérieure est d'environ 650 Da). Dans le cas des porines OmpC et OmpF, le diamètre apparent du pore est différent. En effet, le pore de OmpF est plus large que le pore formé au niveau de OmpC (0,58 nm comparé à 0,54 nm) et la vitesse de diffusion de molécules importantes comme le glucose au travers de OmpF est plus élevée que pour OmpC (Nikaido et Rosenberg, 1983).

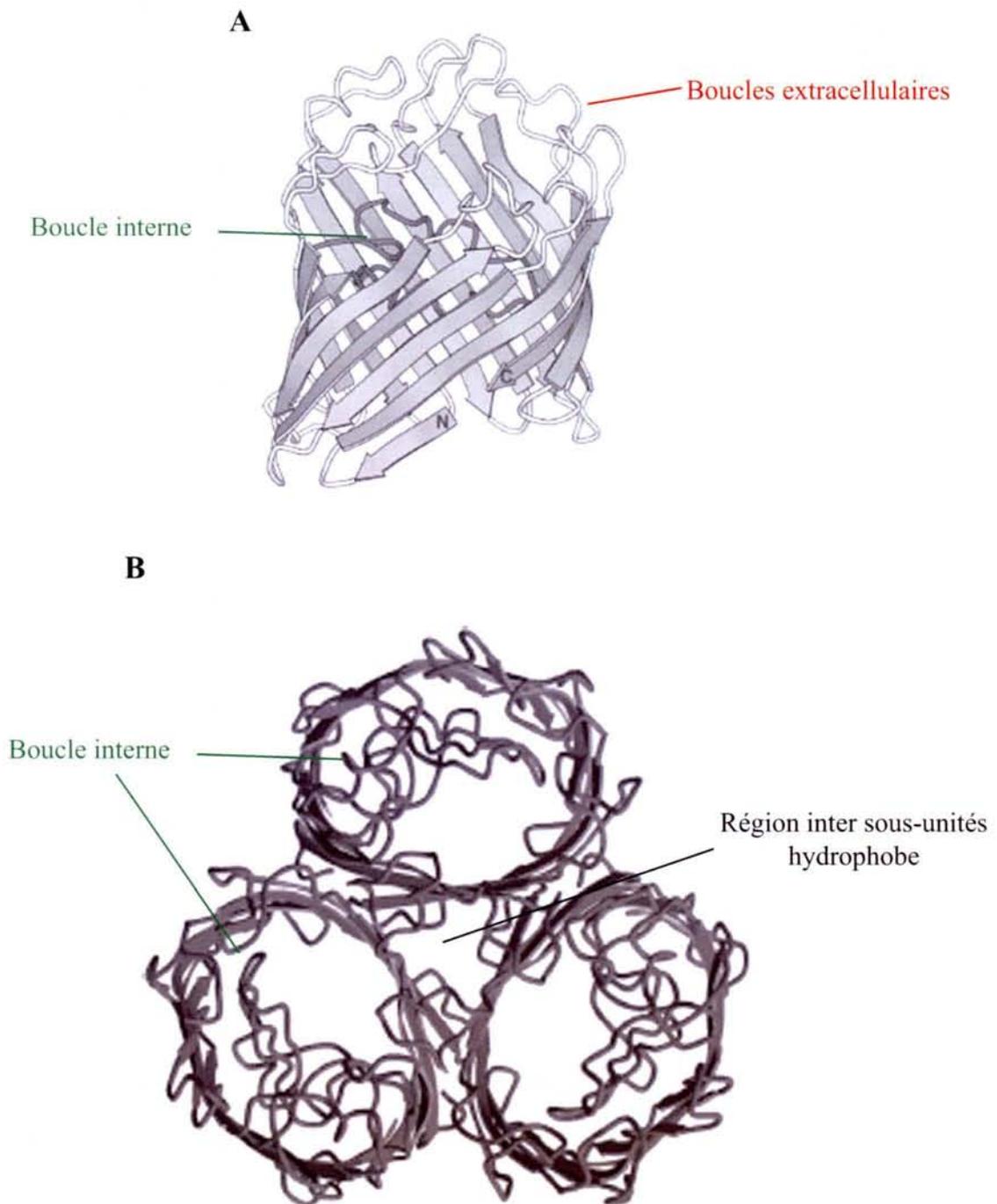


Figure 25 : Structure de la porine OmpF d'*E. coli* (d'après Schirmer, 1998).

A) Structure en tonneau de brins β antiparallèles d'un monomère de la porine OmpF. Le tonneau est composé de 16 brins β , qui sont connectés par de courtes boucles du côté périsplasmique et par des boucles plus longues du côté extracellulaire. La boucle interne qui bloque l'entrée du pore est représentée en gris foncé. Les extrémités N et C terminales sont indiquées par N et C.

B) Représentation schématique d'un trimère de la protéine OmpF. La face représentée est exposée au milieu extracellulaire.

Les porines OmpC et OmpF ont tout de même une sélectivité pour les cations, alors que la porine PhoE montre une sélectivité pour les anions (Nikaido et Rosenberg, 1983). La nature des acides aminés qui constituent la boucle interne est sans doute responsable de cette sélectivité.

Le taux des porines OmpC et OmpF varie en fonction de l'osmolarité du milieu. OmpF est synthétisée préférentiellement à faible osmolarité, alors que OmpC est exprimée à forte osmolarité du milieu (Hall et Silhavy, 1981). Le mécanisme de cette expression différentielle qui est basé sur un système à deux composants sera détaillé dans la partie suivante. La porine PhoE est exprimée uniquement en condition de limitation en phosphate. En effet, sa sélectivité pour les anions permettrait un transport facilité des composés phosphorylés (Korteland *et al.*, 1982).

D'autres protéines sont présentes au niveau de la membrane externe notamment OmpA. Ces protéines auraient de nombreux rôles, notamment elles permettraient de stabiliser la membrane externe. Elles participeraient aussi à la conjugaison bactérienne et pourraient servir de récepteur pour les phages (pour revue : Nikaido, 1996). OmpA pourrait servir de porine mais ce rôle a été remis en cause lors de l'étude de la structure de cette protéine (Pautsch et Schulz, 1998).

1.2) Les porines spécifiques.

Ces porines auraient une spécificité restreinte à certains substrats. L'exemple qui a été le plus étudié est la maltoporine LamB, qui permet le passage des maltodextrines au travers de la membrane externe (Koebnik *et al.*, 2000). Du fait de leur taille, les maltodextrines ne peuvent pas traverser la membrane externe par les porines majeures.

Comme pour les porines non spécifiques, la maltoporine LamB est constituée par l'assemblage de trois monomères. Chacun de ces monomères possède une structure en tonneau de 18 brins β antiparallèles (Schirmer *et al.*, 1995). De plus, chaque monomère a une affinité pour le maltose et les maltodextrines. Le mécanisme de transport des maltodextrines au travers de LamB est complexe et fait appel à des acides aminés aromatiques sur une face du pore, alors que l'autre face est composée d'acides aminés chargés. La translocation du substrat se ferait par formation puis disruption des liaisons hydrogènes entre les maltodextrines et les acides aminés chargés (Meyer et Schulz, 1997). Dans le périplasme, les maltodextrines sont prises en charge par un transporteur ABC spécifique du maltose et

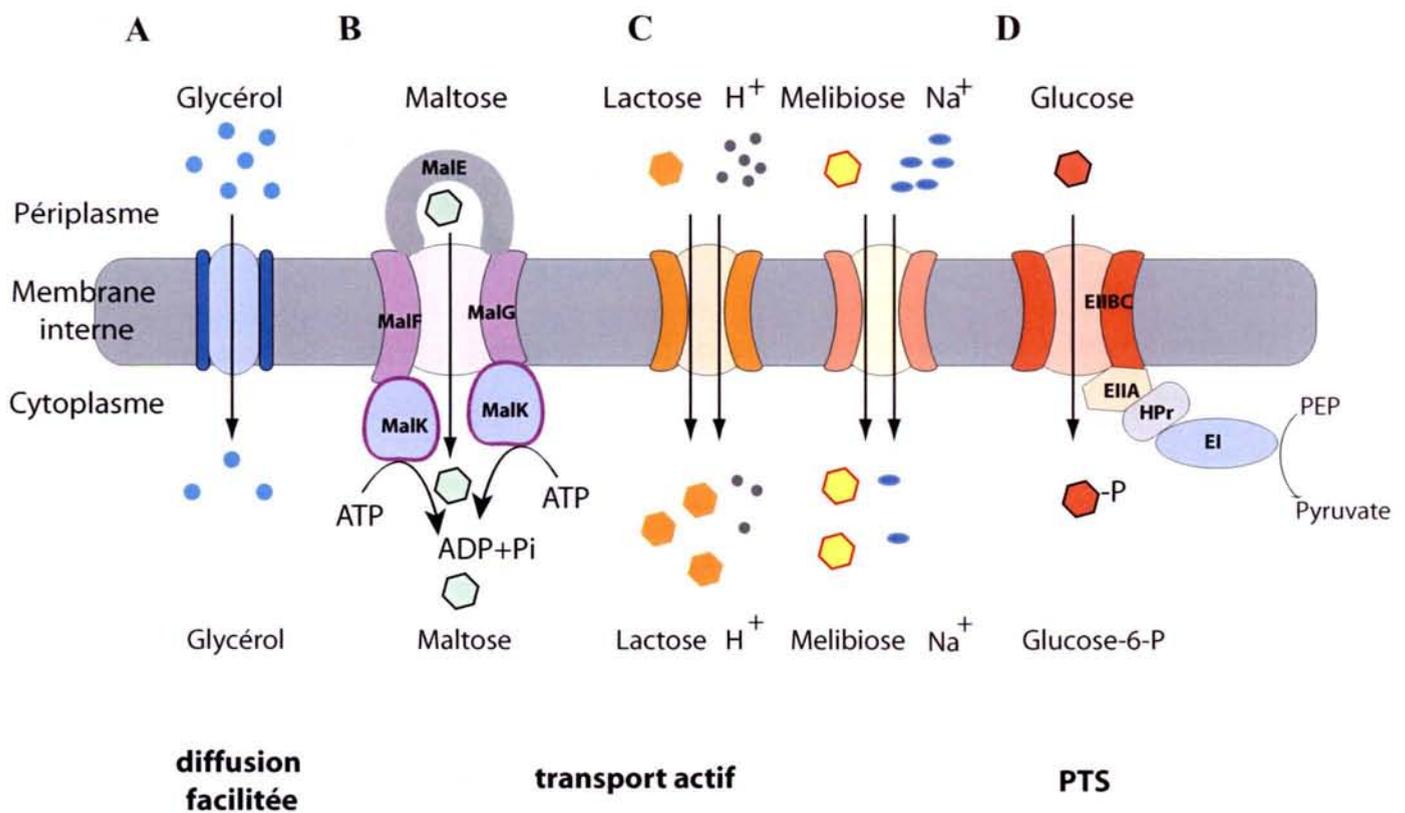


Figure 26 : Organisation des différents systèmes de transport chez *E. coli*.

A) Système de diffusion facilitée. Exemple du glycérol.

B) Transport actif par l'intermédiaire du transporteur ABC du maltose. Dans ce mécanisme de transport, l'ATP est utilisé comme source d'énergie pour incorporer le maltose dans la cellule (pour revue Boos *et al.*, 1998).

C) Transport actif utilisant l'énergie créée par un gradient soit de protons dans le cas du transport du lactose, soit d'ions sodium dans le cas du transport du mélibiose.

D) Transport du glucose par le système PTS. Le glucose est incorporé dans la cellule et phosphorylé par les composants du système PTS. L'énergie nécessaire au transport est apportée par le phosphoénol pyruvate (PEP).

notamment par la protéine périplasmique MalE et les protéines membranaires MalF et MalG (pour revue : Boos et Shuman, 1998).

Il existe aussi des récepteurs de haute affinité au niveau de la membrane externe qui permettent le transport de substrats de grande taille comme, par exemple, les complexes fer-sidérophores (FhuA), la vitamine B12 (BtuB) ou le citrate de fer (FecA) (pour revue : Nikaido, 1996).

D'autres protéines sont retrouvées au niveau de la membrane externe. Ces protéines, par exemple OmpX, permettent la neutralisation des systèmes de défense de la cellule eucaryote dans le cas d'une infection bactérienne (Vogt et Schulz, 1999). Il existe aussi des protéines à activité enzymatique comme la phospholipase A ou OmpT, qui possèdent une activité protéase et seraient impliquées dans l'infection de cellules eucaryotes par les bactéries (Brandenburg *et al.*, 2004).

L'expression de la plupart de ces protéines à activité enzymatique ou ayant une spécificité de transport est induite en fonction des conditions environnementales, alors que les porines non spécifiques sont en général exprimées de façon constitutive, même si des variations dans les quantités sont observables.

Lorsque les nutriments franchissent la membrane externe pour se situer dans le périplasma grâce aux porines et aux autres systèmes de transport que nous venons de décrire, ils sont pris en charge par d'autres transporteurs spécifiques qui se situent au niveau de la membrane interne. Ces transporteurs utilisent différents mécanismes pour permettre aux nutriments de franchir la membrane interne (Fig. 26).

Le premier mécanisme est la diffusion facilitée qui permet, par exemple, le passage du glycérol à travers la membrane interne. Les autres mécanismes utilisent de l'énergie pour fonctionner. Il s'agit par exemple des transporteurs ABC qui utilisent de l'ATP et qui permettent, par exemple, l'incorporation du maltose dans la cellule (pour revue : Boos et Shuman, 1998). D'autres substrats comme le lactose ou le mélibiose sont incorporés grâce à des symporteurs (Wilson et Ding, 2001; Abramson *et al.*, 2003). Ces transporteurs permettent l'incorporation du substrat en même temps que l'incorporation d'ions H^+ ou Na^+ . L'énergie utilisée est créée par le gradient ionique formé entre les deux compartiments.

Le mécanisme utilisé pour l'incorporation des glucosides est basé sur la phosphorylation de protéines. Il s'agit des systèmes PTS qui sont ubiquitaires chez les bactéries et qui permettent à la fois le transport et la phosphorylation de la source carbonée. Je vais maintenant détailler ces systèmes de transport.

II.) Le système PTS chez *E. coli*.

Le système PTS qui permet le transport à l'intérieur de la cellule et la phosphorylation des composés osidiques est un système de phospho-relais. En effet, le phosphate qui est issu du phosphoénol pyruvate est transmis au substrat qui entre dans la cellule par l'intermédiaire des protéines EI, HPr et EII (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996).

Les protéines EI et HPr sont communes aux différents systèmes PTS, alors que les protéines EII sont spécifiques de la source de carbone qui est transportée.

La protéine EI peut s'autophosphoryler en utilisant le phosphoénol pyruvate (PEP) qui est un intermédiaire de la glycolyse, comme donneur de phosphate (Fig. 27). La protéine EI transfère ensuite son phosphate à la protéine HPr. A ce niveau, HPr peut transmettre son phosphate aux différentes protéines EII spécifiques des différents composés osidiques. Ces protéines sont en général composées de trois domaines structuraux (A, B et C) qui peuvent être associés de manière différente suivant les systèmes EII (Fig. 27). La protéine EI_A est phosphorylée par HPr, puis elle transmet son phosphate au domaine EI_B, qui phosphoryle le substrat. Ces deux domaines sont cytoplasmiques, alors que le domaine EI_C, qui est membranaire, permet le passage de la source de carbone au travers de la membrane interne (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996; Siebold *et al.*, 2001).

La phosphorylation des protéines EI, HPr et EII est réversible, alors que la phosphorylation du composé osidique est irréversible. La phosphorylation de EI entraîne la transformation du PEP en pyruvate. Dans le cas du système PTS, cette phosphorylation est réversible alors que les enzymes de la glycolyse qui réalisent cette réaction (pyruvate kinase) ont un mécanisme irréversible. Le système de transport et de phosphorylation des sucres est donc dépendant du métabolisme cellulaire.

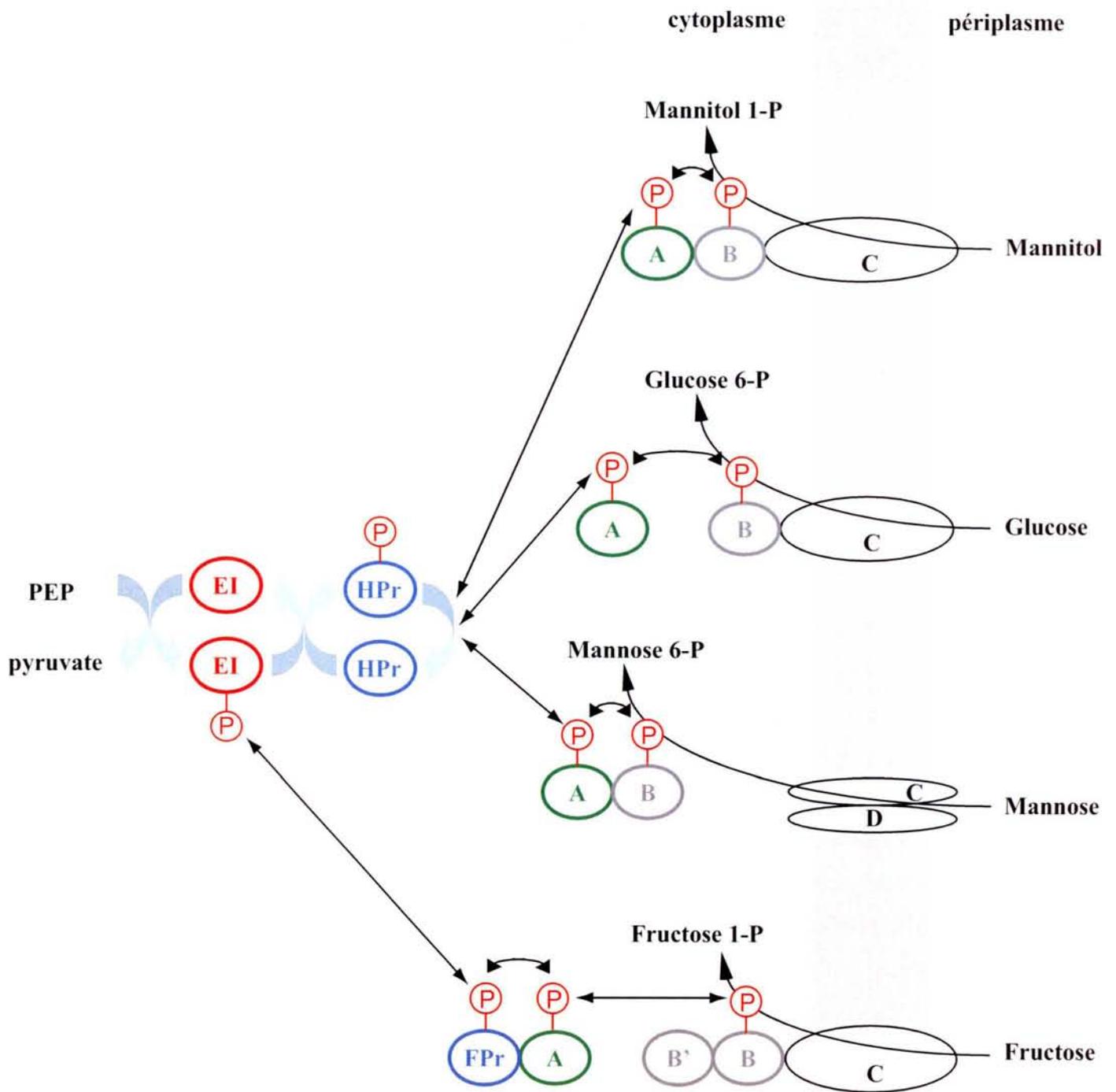


Figure 27 : Organisation du système PTS du mannitol, du glucose, du mannose et du fructose chez *E. coli* (pour revue : Postma *et al.*, 1994 ; Postma *et al.*, 1996).

Les protéines générales EI et HPr, ainsi que le domaine de la protéine FPr du système PTS du fructose, homologue à la protéine HPr, sont représentés en rouge et en bleu respectivement. Les domaines A des enzymes EII spécifiques des différents carbohydrates sont schématisés en vert, tandis que les domaines B de ces enzymes sont représentés en gris et les domaines C et D transmembranaires en noir. Le groupement phosphate, qui est transféré depuis le PEP jusqu'à la source de carbone, est représenté par un P cerclé en rouge.

II.1) Organisation protéique et génétique des composants du système PTS.

II.1.1) Organisation protéique.

II.1.1.1) La protéine EI.

L'enzyme I (64 kDa) est constituée de deux domaines : un domaine N-terminal compact qui contient le site actif et un domaine C-terminal dont la structure est beaucoup plus flexible et qui serait impliqué dans la dimérisation de la protéine.

EI peut s'autophosphoryler sur chaque monomère, ceci au niveau d'un résidu histidine du domaine N-terminal (His 189), et en utilisant le PEP comme donneur de phosphate en présence de Mg^{2+} (pour revue : Ginsburg et Peterkofsky, 2002). Cette phosphorylation n'est réalisée qu'avec la forme dimérique de EI (Seok *et al.*, 1998).

De plus, la phosphorylation de l'histidine entraîne sa rotation vers la surface de la protéine, ce qui entraîne une déstabilisation du domaine N-terminal et augmente la capacité de dimérisation de EI (Ginsburg et Peterkofsky, 2002). La forme dimérique serait donc stabilisée. Après le transfert du phosphate à HPr, l'équilibre entre les formes dimériques et monomériques serait rétabli jusqu'à ce que EI soit à nouveau phosphorylée par le PEP.

Le taux de phosphorylation de EI est donc régulé par le taux de PEP dans la cellule, ce qui lie le transport des sucres à leur métabolisme.

II.1.1.2) La protéine HPr.

Le phosphate de EI est ensuite transféré à la protéine monomérique HPr (Histidine phosphocarrier protein 9 kDa) qui est phosphorylée sur une histidine (His 15). La structure 3D de HPr en interaction avec la protéine EI (domaine N-terminal) a été déterminée par RMN (Fig. 28) (Garrett *et al.*, 1999). Elle montre que lors de l'interaction, la structure des deux protéines change peu. La spécificité du complexe dépend d'interactions de surface entre les deux protéines. Le phosphate stabilise la structure de la protéine HPr, ce qui explique le transfert du phosphate de EI vers HPr (Garrett *et al.*, 1999). HPr phosphoryle un grand nombre de cibles dont les structures tridimensionnelles sont très différentes (Siebold *et al.*, 2001). La structure globale de HPr ne change pas au cours des interactions. De plus, la zone de HPr qui permet la reconnaissance des différents partenaires est la même. Ainsi, la

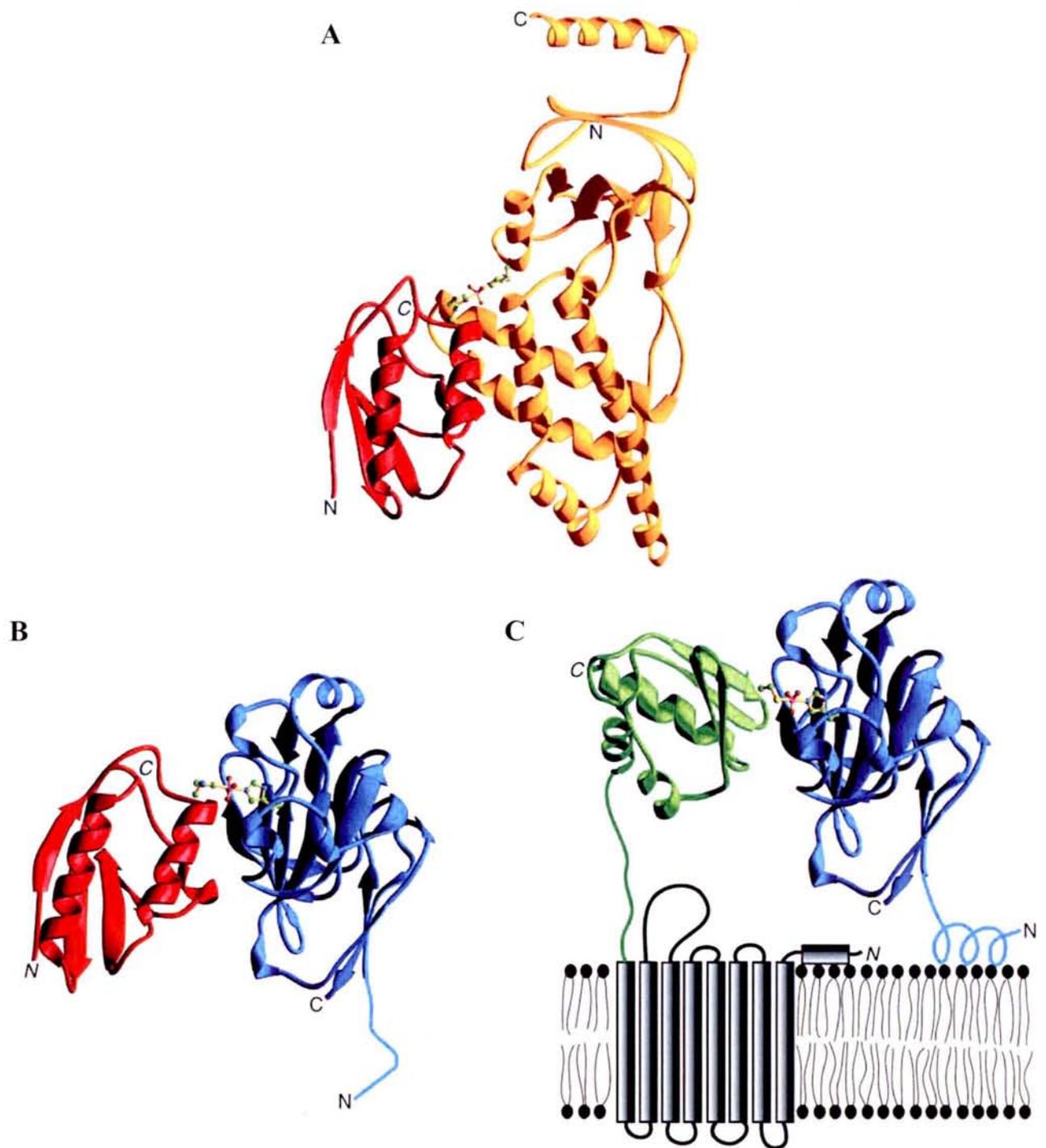


Figure 28 : Structure des protéines du système PTS du glucose d'*E. coli* (d'après Cai *et al.*, 2003). Les différentes protéines qui permettent le transfert du phosphate du PEP au glucose sont représentées en orange pour EI, en rouge pour HPr, en bleu pour EIIA^{Glc}, en vert pour EIIB^{Glc} et en gris pour EIIC^{Glc}. Les histidines qui permettent le transfert du phosphate sont représentées en jaune. L'état de transition potentiel du phosphate au cours du transfert est représenté. Les extrémités N et C terminales sont indiquées pour chaque protéine.

A) Interaction entre EI et HPr.
 B) Interaction entre HPr et EIIA^{Glc}.
 C) Interaction entre EIIA^{Glc} et EIIB^{Glc}. EIIC^{Glc} est représenté par 8 hélices transmembranaires. Dans la structure de EIIA^{Glc} en solution, l'extrémité N-terminale n'est pas structurée, alors que lors de l'interaction avec EIIB^{Glc}, cette hélice se structure pour former une hélice amphipatique qui permet un ancrage de EIIA^{Glc} au niveau de la membrane et stabilise l'interaction entre ces deux protéines.

spécificité de l'interaction ne serait due qu'à une modification de l'orientation des chaînes latérales d'acides aminés en fonction du partenaire (Cai *et al.*, 2003).

Le système PTS spécifique du fructose présente la particularité d'associer un homologue de HPr et la protéine EIIA spécifique du fructose (Fig. 27). Cette protéine appelée FPr peut être directement phosphorylée par EI, transmettre son phosphate au domaine N-terminal (protéine EIIA^{Fru}) et permettre l'incorporation du fructose via le transporteur EIIBC^{Fru}. FPr pourrait aussi phosphoryler les protéines EIIA d'autres systèmes PTS mais avec une efficacité moindre que HPr. Cette réaction pourrait permettre la croissance de souches mutées dans le gène codant la protéine HPr (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996).

II.1.1.3) Les complexes EII spécifiques et le mécanisme de translocation.

Ces complexes sont constitués principalement de trois domaines A, B, C, mais un quatrième domaine EIID est présent au niveau de la membrane dans certains systèmes PTS comme le transporteur du mannose (Fig. 27). L'analyse du génome d'*E. coli* a permis l'identification de 22 protéines EII spécifiques. Les substrats de 13 d'entre elles sont connus actuellement (Plumbridge, 2002). L'organisation de ces domaines varie en fonction du système PTS mais aussi en fonction de la bactérie (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996; Siebold *et al.*, 2001). Dans le cas du système PTS spécifique du glucose, le domaine EIIA^{Glc} est cytoplasmique, tandis que le domaine EIIB cytoplasmique est fusionné au domaine membranaire EIIC. Dans le cas du système PTS spécifique du mannose, les domaines EIIA et EIIB sont fusionnés et se trouvent dans le cytoplasme, alors que dans le cas du système de transport du mannitol, les deux domaines A et B sont fusionnés au domaine membranaire C (Fig. 27). Il est important de noter que même si les transporteurs sont spécifiques d'un sucre, ils peuvent transporter d'autres hexoses avec en général des affinités plus faibles. Par exemple, le glucose peut être transporté à la fois par son transporteur spécifique EIIA-EIIBC^{Glc} mais aussi par le transporteur spécifique du mannose EIAB-EIICD^{Man}. Par ailleurs, certains domaines EIIA peuvent phosphoryler d'autres cibles que les domaines EIIB qui font partie du même système PTS. Ainsi, EIIA^{Glc} peut phosphoryler le domaine EIIB spécifique du tréhalose (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996).

Les domaines A et B sont phosphorylés au niveau d'une histidine ou d'une cystéine pour certains domaines EIIB. Au niveau structural, la structure 3D de HPr en interaction avec la protéine EIIA spécifique du glucose a été déterminée par RMN (Fig. 28 B) (Wang *et al.*,

2000). Les surfaces d'interaction de HPr avec EI et de HPr avec EIIA^{Glc} sont les mêmes, ce qui indique que le transfert du phosphate est séquentiel. En effet, le complexe entre EI et HPr doit se dissocier pour que HPr puisse interagir avec EIIA^{Glc} (Wang *et al.*, 2000). Par ailleurs, la structure des domaines EIIA et EIIB est très différente suivant le système PTS (Siebold *et al.*, 2001).

Dans le cas du système PTS spécifique du glucose, EIIB^{Glc}, qui est fusionné au domaine EIIC^{Glc}, est phosphorylé au niveau d'une cystéine (Cys 421, (Nuoffer *et al.*, 1988)) par EIIA^{Glc} (qui est phosphorylé au niveau de His 90). Dans la membrane, cette protéine EIIBC^{Glc} est dimérique (Lanz et Erni, 1998). Le domaine EIIC, constitué de 8 hélices transmembranaires, est relié au domaine EIIB^{Glc} par une chaîne flexible (Fig. 28 C).

Il est à noter que les surfaces d'interaction entre EIIA^{Glc}-HPr et EIIA^{Glc}-EIIB^{Glc} sont les mêmes, ce qui implique, comme dans le cas du transfert du phosphate entre EI et HPr, que l'échange du phosphate est séquentiel (Fig. 28 C) (Cai *et al.*, 2003).

Le transfert du phosphate repose donc sur une série de réactions bimoléculaires entre les composants du PTS. Son caractère directionnel est lié à une plus grande stabilité des phosphorylations portées par les protéines au fur et à mesure du transfert.

Le phosphate est ensuite transféré au glucose par un mécanisme encore inconnu. Le domaine EIIC^{Glc} constitue le domaine de fixation du glucose, ce qui a été démontré par l'étude de mutations qui modifiaient la spécificité de reconnaissance du substrat (Zeppenfeld *et al.*, 2000; Notley-McRobb et Ferenci, 2000; Kornberg *et al.*, 2000).

Le transport du glucose au travers de la membrane serait plus ou moins indépendant de sa phosphorylation par EIIB^{Glc}. La phosphorylation pourrait jouer deux rôles dans la fonction des protéines EII (pour revue : Robillard et Broos, 1999) :

- 1) permettre le mouvement du sucre depuis le site de fixation périplasmique vers le site cytoplasmique (transport).
- 2) permettre le relargage de l'hexose-phosphate dans le cytoplasme. En effet, les protéines EII peuvent catalyser la phosphorylation d'un hexose déjà présent dans le cytoplasme.

Certaines mutations de la protéine EIIBC^{Glc} spécifique du glucose permettent de découpler les deux activités de transport et de phosphorylation (Buhr *et al.*, 1992; Ruijter *et al.*, 1992).

Un modèle permettant d'expliquer un grand nombre d'observations a été mis au point par l'équipe de B. Erni (Garcia-Alles *et al.*, 2002a). EIIC^{Glc} aurait un site de faible affinité et un site de forte affinité pour le substrat au niveau cytoplasmique, alors qu'il n'y aurait qu'un seul site de forte affinité au niveau du périplasma (il pourrait y avoir un deuxième site de faible affinité du fait de la symétrie entre les deux monomères de EIIBC^{Glc}). Le substrat qui se situe

au niveau du périplasme serait pris en charge par le site de forte affinité qui ferait partie de la même sous-unité que le site de faible affinité cytoplasmique. Le substrat serait transféré entre ces deux sites en même temps qu'il serait phosphorylé par EIIB^{Glc}. L'autre site de forte affinité cytoplasmique serait responsable de la phosphorylation intracellulaire de différents substrats (par EIIB^{Glc}) (Garcia-Alles *et al.*, 2002a; Garcia-Alles *et al.*, 2002b).

II.1.2) Organisation des gènes du système PTS.

En général, les gènes codant les enzymes EII spécifiques d'une source de carbone sont situés au même *locus* que les gènes qui contrôlent l'expression des protéines du catabolisme de cette source carbonée. Cet ensemble de gènes se trouve en général sous forme d'un opéron fortement régulé par la présence de la source de carbone à l'extérieur de la cellule. Les différents opérons qui permettent le métabolisme d'un grand nombre de sources carbonées sont répartis le long du chromosome d'*E. coli* (Postma *et al.*, 1993; Postma *et al.*, 1996). Les protéines EI et HPr communes à tous les systèmes PTS sont codées par les gènes *ptsI* et *ptsH* qui forment un opéron appelé *pts* avec le gène *crr* codant la protéine EIIA^{Glc} spécifique du glucose. Le gène *ptsG* codant le transporteur spécifique du glucose EIIBC^{Glc} est localisé à une autre position sur le chromosome (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996). La régulation de l'expression de ces gènes sera décrite dans le chapitre II.2.2.

II.2) Régulations exercées par le système PTS.

Les sources de carbone transportées par l'intermédiaire d'un système PTS sont utilisées préférentiellement par la cellule au détriment d'autres sources de carbone transportées par un autre moyen. La diauxie décrite par Monod repose sur l'utilisation préférentielle d'un sucre A ou sucre PTS (comme le glucose) par rapport à un sucre B ou sucre non PTS (comme le lactose). Dans un premier temps, le glucose est consommé, alors que le lactose ne sera consommé qu'après épuisement du glucose. En présence de glucose (ou d'un autre sucre transporté par le système PTS), l'expression des gènes du métabolisme du lactose est réprimée par deux mécanismes différents.

Le premier mécanisme est la limitation de l'entrée de l'inducteur dans la cellule, ce qui permet de lever la répression par la protéine LacI au niveau du promoteur de l'opéron lactose. Ce phénomène qui peut être généralisé pour un grand nombre d'autres opérons réprimés par des répresseurs spécifiques a été appelé exclusion de l'inducteur.

Le deuxième mécanisme est basé sur le fait qu'une forte initiation de la transcription au niveau du promoteur *lac* nécessite la fixation du complexe CRP-AMPc. La présence d'un sucre transporté par le système PTS entraîne une diminution de la concentration intracellulaire de ce complexe, ce qui limite l'expression de l'opéron lactose mais aussi d'autres opérons régulés par ce complexe. Ce phénomène général a été appelé répression catabolique.

Ces deux phénomènes sont liés au taux de phosphorylation de la protéine EIIA^{Glc}, ce qui indique le rôle particulier que joue cette protéine dans la régulation du métabolisme des sources de carbone. De plus, le taux de phosphorylation des autres composants du système PTS spécifique du glucose peuvent aussi jouer un rôle dans ces régulations. En effet, lorsqu'un sucre PTS autre que le glucose est transporté par son système EIIABC spécifique, il en résulte une déphosphorylation de HPr. Sachant que la phosphorylation des protéines du système PTS est réversible, EIIA^{Glc} va pouvoir transférer son phosphate à HPr. Le taux de phosphorylation de EIIA^{Glc} va donc aussi varier en présence d'un autre sucre PTS que le glucose.

D'autre part, la synthèse des composants du système PTS est très fortement régulée pour répondre aux conditions environnementales. Ce contrôle s'effectue majoritairement au niveau transcriptionnel comme nous le verrons plus loin.

II.2.1) Régulations par les composants du système PTS.

II.2.1.1) Régulation exercée par EI.

La protéine EI sous forme non phosphorylée et monomérique permet d'inhiber l'autophosphorylation de la protéine membranaire CheA qui est responsable de la régulation de l'activité du flagelle chez *E. coli* (Lux *et al.*, 1999). EI est majoritairement sous forme non phosphorylée lors du transport d'un sucre PTS comme le glucose. La bactérie se déplace donc vers le gradient de concentration croissant de sucre PTS (Lux *et al.*, 1999).

II.2.1.2) Régulation exercée par EIIA^{Glc}.

II.2.1.2.1) Phénomène d'exclusion de l'inducteur (Fig. 29).

L'exclusion de l'inducteur est spécifique à la présence de sucres non PTS (B) dans le milieu extérieur comme par exemple le lactose, le maltose, le glycérol et le mélibiose en même

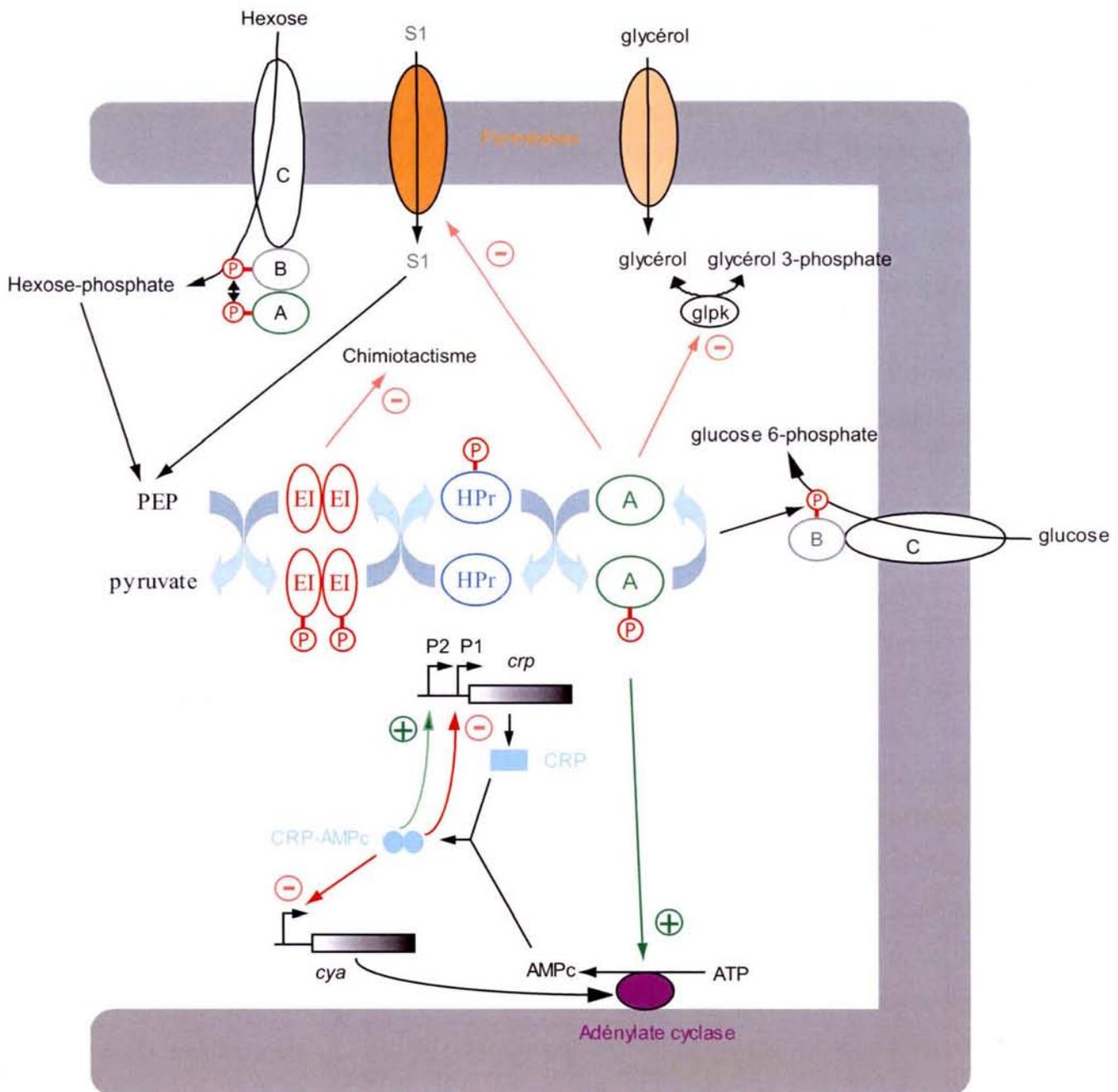


Figure 29 : Régulations dépendant du système PTS spécifique du glucose chez *E. coli* (d'après Postma *et al.*, 1993 ; Postma *et al.*, 1996 ; Ishizuka *et al.*, 1994).

Les protéines générales EI et HPr sont représentées respectivement, en rouge et en bleu. Le domaine EIIGlc est schématisé en vert, tandis que EIIBGlc est représenté en gris et EIICGlc en noir. Le groupement phosphate qui est transféré depuis le PEP jusqu'à la source de carbone est représenté par un P encerclé en rouge. Les perméases spécifiques pour le glycérol et les sucres S1 (lactose ou mélibiose) sont représentées en orange. La glycérol kinase et l'adénylate cyclase sont représentées en noir et en violet, respectivement. Les gènes *cya* et *crp* sont indiqués, ainsi que les promoteurs en amont de ces gènes. Le complexe CRP-AMPC est indiqué en bleu. Les régulations négatives sont indiquées par des flèches rouges et les régulations positives par des flèches vertes. L'entrée des sucres S1 ou des hexoses phosphate par leurs systèmes de transport respectifs (perméase ou autre système PTS) et le métabolisme permet la formation du PEP par la voie de la glycolyse.

temps qu'un sucre PTS (A) comme par exemple le glucose, le fructose et le mannose (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996).

Lorsqu'un sucre PTS comme le glucose est présent dans le milieu périplasmique, le phosphate issu du PEP est transféré aux protéines EI et HPr, puis à EIIA^{Glc} qui le transmet à EIIBC^{Glc} puis au glucose. Il en résulte que le taux de ces protéines sous leur forme phosphorylée diminue.

La protéine EIIA^{Glc} sous sa forme déphosphorylée peut interagir directement avec certaines protéines impliquées dans le transport et le métabolisme des sources de carbone non PTS (Fig. 29). Il s'agit par exemple des perméases du lactose (LacY), du mélibiose (MelB), du raffinose (RafB) et du maltose (MalK) mais aussi de la glycérol kinase (GK), qui catalyse la première étape du métabolisme du glycérol (Saier *et al.*, 1978; Wilson *et al.*, 1990; Kuroda *et al.*, 1992; Titgemeyer *et al.*, 1994). Il est à noter que la phosphorylation de EIIA^{Glc} empêche son interaction avec ces différentes protéines.

Cette interaction directe, qui a été démontrée *in vitro* dans le cas des perméases du lactose et du maltose et de la glycérol kinase, limite la quantité intracellulaire de l'inducteur des voies métaboliques alternatives (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996). En effet, la fixation de EIIA^{Glc} entraîne une inhibition du transport des sucres dans le cas des perméases et une diminution de l'activité enzymatique dans le cas de la glycérol kinase.

Le site de fixation de EIIA^{Glc} sur la glycérol kinase n'est pas situé à proximité du site actif de l'enzyme, ce qui implique une modification allostérique pour qu'il y ait inhibition de son activité (pour revue : Postma *et al.*, 1996; Postma *et al.*, 1993). La production de glycérol 3-phosphate, qui est l'inducteur des gènes du catabolisme du glycérol, est donc inhibée. Comme le glycérol pénètre dans la cellule par diffusion facilitée, l'inhibition de l'activité de la glycérol kinase entraîne une réduction de l'entrée du glycérol (Bruckner et Titgemeyer, 2002). Dans le cas de la perméation du lactose, la fixation de EIIA^{Glc} sur la perméase empêche l'entrée du lactose dans la cellule. Il ne sera donc pas métabolisé pour former l'allolactose qui est l'inducteur de l'opéron lactose. Cet inducteur permet en effet la levée de répression de la protéine LacI fixée au niveau du promoteur du gène *lac* et empêche ainsi la synthèse des enzymes impliquées dans le catabolisme du lactose. Il en résulte une faible expression de cet opéron lorsque du glucose est présent.

De plus, l'interaction entre EIIA^{Glc} et la perméase du lactose n'est possible que lorsque le substrat est présent (Saier *et al.*, 1983; Seok *et al.*, 1997b). Lors de l'interaction du substrat avec LacY, il y aurait un changement de structure de la perméase qui permettrait l'interaction avec EIIA^{Glc} (Seok *et al.*, 1997b). Ce mécanisme permettrait d'éviter la fixation inutile de

EIIA^{Glc} sur LacY en absence de substrat et ainsi la titration par l'ensemble des protéines cibles de cette protéine.

Par mutagenèse du gène codant LacY et en étudiant les propriétés de fixation de EIIA^{Glc} sur les mutants produits, Sondej *et al.* (1999) ont défini la séquence de reconnaissance de EIIA^{Glc}. Cette séquence, qui se situe au niveau des boucles cytoplasmiques de LacY, est composée de deux régions d'interaction. Le consensus de la région I est (V, L)-0, 2-(S, A, T)-0, 2-(R)-0, 6-(G)(K, R) et (S, T)-0, 4-(V, L)-0, 2-(N, Q, E)-0, 2-(G)-0, 1-(N, K)-0, 1-(S) pour la région II (Sondej *et al.*, 1999) (entre parenthèse : acides aminés trouvés préférentiellement et entre tirets : nombre d'acides aminés qui séparent ces séquences). De plus, les auteurs ont émis l'hypothèse que lors de la fixation du substrat sur LacY, le changement de conformation permet le rapprochement de ces deux séquences, ce qui facilite l'interaction avec EIIA^{Glc}. Cette étude a été poursuivie en recherchant des protéines qui possèdent ces deux séquences. En effet, ces protéines sont susceptibles d'interagir avec EIIA^{Glc}. Les auteurs ont mis en évidence 37 protéines qui possédaient ces deux régions (Sondej *et al.*, 1999). Ils ont ainsi retrouvé les perméases du lactose et du maltose, ainsi que du raffinose. D'autres protéines avec des activités variées possèdent cette séquence consensus. Ainsi, des protéines qui seraient impliquées dans la signalisation comme la protéine PhoQ, qui fait partie du système à deux composants PhoP/Q, pourraient fixer EIIA^{Glc}. Les autres protéines connues pour interagir avec EIIA^{Glc} n'ont pas été détectées par cette approche. En effet, la perméase du mélibiose ne contient que la région consensus I. Ainsi, le nombre des protéines pouvant fixer EIIA^{Glc} serait beaucoup plus élevé qu'attendu jusque là (jusqu'à 1400 protéines) (Sondej *et al.*, 1999). EIIA^{Glc} serait donc un régulateur général.

La zone de EIIA^{Glc} qui entre en interaction avec HPr et EIIBC^{Glc} est la même que celle utilisée pour interagir avec les perméases et la glycérol kinase (Wang *et al.*, 2000; Cai *et al.*, 2003). L'histidine phosphorylée est aussi située dans cette zone. Ainsi, la phosphorylation empêche l'interaction de EIIA^{Glc} avec ses protéines cibles.

D'autres sources de carbone PTS permettent aussi la répression de l'expression des gènes du catabolisme des sources de carbone alternatives (non PTS). Cette répression est basée sur l'interaction de la protéine EIIA^{Glc} avec les perméases permettant le transport de ces sources de carbone. En effet, la phosphorylation entre EIIA^{Glc} et HPr est réversible. Ainsi, si un sucre PTS est présent dans le milieu, HPr va transmettre son phosphate au transporteur spécifique EIIA de ce sucre PTS, ce qui va entraîner la déphosphorylation de EIIA^{Glc}. Cette protéine inhiberait alors les protéines cibles du phénomène d'exclusion de l'inducteur (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996).

De façon surprenante, certaines sources de carbone non PTS permettent une exclusion de l'inducteur. Ce phénomène qui a été observé avec le lactose, le mélibiose et le glucose 6-phosphate serait lié au métabolisme de ces sources de carbone (Hogema *et al.*, 1997; Hogema *et al.*, 1998b; Hogema *et al.*, 1998a; Hogema *et al.*, 1999). La cible de cette régulation serait EIIA^{Glc} qui inhibe l'entrée des inducteurs dans la cellule. Le métabolisme des sources de carbone non PTS pourrait modifier le rapport PEP/pyruvate, ce qui pourrait modifier le taux de phosphorylation de EIIA^{Glc}, ainsi que des autres protéines du système PTS (Hogema *et al.*, 1998a).

II.2.1.2.2) Modulation de l'activité de l'adénylate cyclase.

En absence de glucose ou d'un sucre PTS dans le milieu de culture, les protéines du système PTS spécifique du glucose sont majoritairement sous la forme phosphorylée. La protéine EIIA^{Glc} sous sa forme phosphorylée (EIIA^{Glc}-P) permet d'activer l'adénylate cyclase, codée par le gène *cya* qui est responsable de la synthèse d'AMPc dans la cellule (Fig. 29) (Reddy et Kamireddi, 1998; Krin *et al.*, 2002 et pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996; Stulke et Hillen, 1999; Bruckner et Titgemeyer, 2002). L'adénylate cyclase est constituée de deux domaines : le domaine N-terminal qui porte l'activité catalytique et le domaine C-terminal qui régule l'activité du domaine N-terminal. Une délétion de la région C-terminale entraîne une absence de régulation par la présence de glucose, ce qui montre que ce domaine est impliqué dans la régulation par EIIA^{Glc} (Crasnier *et al.*, 1994). Du fait de l'activation de l'adénylate cyclase par EIIA^{Glc}-P, le taux d'AMPc augmente dans la cellule en absence de glucose. L'AMPc se fixe sur la protéine CRP pour former le complexe CRP-AMPc (cf. partie 1, paragraphe II.3.2.1.2) et pour revue : Botsford et Harman, 1992). L'augmentation de la concentration en ce complexe permet d'activer la transcription d'un grand nombre de gènes et d'opérons qui contrôlent l'expression de protéines et d'enzymes impliquées dans le métabolisme des sources de carbone alternatives (pour revue : Botsford et Harman, 1992; Postma *et al.*, 1993; Postma *et al.*, 1996; Saier *et al.*, 1996; Stulke et Hillen, 1999; Bruckner et Titgemeyer, 2002). CRP-AMPc régule aussi l'expression d'autres facteurs plus généraux comme la protéine Fis (Nasser *et al.*, 2001) ou le régulateur transcriptionnel FlhD₂C₂ qui contrôle l'expression des flagelles (Soutourina *et al.*, 1999).

En présence d'un sucre PTS, le taux d'AMPc diminue du fait de l'absence d'activation de l'adénylate cyclase par EIIA^{Glc}-P qui est majoritairement sous forme déphosphorylée. En conséquence, le taux de CRP-AMPc diminue, ce qui limite l'expression des opérons du

métabolisme des sources carbonées alternatives. C'est cet effet qui a été appelé répression catabolique.

Récemment, une approche par puces à ADN a été utilisée pour étudier de manière globale le rôle du complexe CRP-AMPc dans l'expression des gènes, en fonction de la source de carbone (Gosset *et al.*, 2004). Les résultats de cette étude montrent que les gènes régulés par CRP-AMPc codent principalement des enzymes des voies du métabolisme du carbone comme le cycle de Krebs, mais aussi des enzymes et des transporteurs qui initient le métabolisme de sources de carbone comme le galactitol ou des acides aminés (Gosset *et al.*, 2004).

Dans le cas de l'opéron lactose, la répression en présence de glucose serait majoritairement due à l'exclusion de l'inducteur plutôt qu'à la diminution du taux de complexe CRP-AMPc (Inada *et al.*, 1996a; Kimata *et al.*, 1997). Les auteurs ont utilisé un promoteur *lac* qui ne nécessite plus CRP-AMPc pour être efficace. Dans ce cas, l'ajout d'IPTG qui permet de lever la répression par LacI, élimine la répression catabolique. La répression catabolique serait donc principalement due à l'exclusion de l'inducteur (Kimata *et al.*, 1997).

De plus, le taux du complexe CRP-AMPc en présence de glucose est légèrement plus élevé qu'en présence de lactose, ce qui n'est pas expliqué par l'activation de l'adénylate cyclase par $EIIA^{Glc}$ -P (Inada *et al.*, 1996a). Ce fait pourrait être expliqué par le rapport PEP/pyruvate qui serait différent entre culture en glucose et culture en lactose (Hogema *et al.*, 1998a). En effet, le métabolisme du lactose pourrait déterminer le taux de phosphorylation de $EIIA^{Glc}$. Si le transport du lactose est très efficace, le rapport PEP/pyruvate et le taux de phosphorylation de $EIIA^{Glc}$ diminuent. Il en résulte une inhibition des perméases du lactose. Ce système de contrôle permet un transport du lactose en rapport avec son métabolisme. Si ce système est inactivé, il en résulte une excrétion de glucose à l'extérieur de la cellule reflétant un état saturé du métabolisme (Hogema *et al.*, 1999).

Par contre, la concentration en AMPc augmente au cours de la transition entre la consommation des deux sucres (Inada *et al.*, 1996a). Le complexe CRP-AMPc permettrait donc l'adaptation rapide pour la consommation du lactose puis sa concentration diminuerait lorsque le lactose est consommé plus efficacement.

Le complexe CRP-AMPc serait tout de même important dans le phénomène de répression catabolique du fait de son effet activateur sur la transcription du gène *ptsG* qui code $EIIBC^{Glc}$. Le transporteur du glucose est en effet responsable de la déphosphorylation de $EIIA^{Glc}$ en présence de glucose, ce qui pourrait renforcer l'exclusion de l'inducteur (Kimata *et al.*, 1997) mais aussi réduire l'activité de l'adénylate cyclase (Takahashi *et al.*, 1998).

Les régulations effectuées par la protéine EIIA^{Glc} sous forme phosphorylée ou non phosphorylée permettent l'adaptation du transport en fonction de la source de carbone et en fonction du métabolisme de ces sources de carbone. Un autre rôle de cette protéine serait de modifier la balance entre respiration et fermentation au niveau métabolisme. En effet, EIIA^{Glc} non phosphorylée peut interagir avec la protéine FrsA (fermentation/respiration switch protein) (Koo *et al.*, 2004). La disruption du gène codant cette protéine entraîne une augmentation de la respiration cellulaire, alors que sa surproduction aboutit à un taux de fermentation plus élevé c'est-à-dire l'accumulation d'acides organiques. EIIA^{Glc} pourrait ainsi réguler les flux métaboliques respiratoires et fermentaires en interagissant avec FrsA (Koo *et al.*, 2004).

II.2.1.3) Modulation de l'activité du répresseur Mlc par EIIBC^{Glc}.

Mlc (making larger colony) est une protéine de 44 kDa qui a été identifiée par Hosono *et al.*, (1995). La surexpression de cette protéine à partir d'un plasmide multicopie permettait une consommation complète du glucose présent dans le milieu de culture (Hosono *et al.*, 1995). Elle existe sous forme de tétramère et régule l'expression des gènes impliqués dans le métabolisme des sources de carbone (pour revue : Plumbridge, 2002). Cette protéine qui appartient à la famille des répresseurs procaryotiques ROK, inhibe l'initiation de la transcription au niveau des gènes *manXYZ* (Plumbridge, 1998a), *malT* (Decker *et al.*, 1998) et *ptsG* (Kimata *et al.*, 1998; Plumbridge, 1998b), ainsi que de l'opéron *pts* (Kim *et al.*, 1999; Plumbridge, 1999; Tanaka *et al.*, 1999) (pour revue : Plumbridge, 2002). Ces cibles de Mlc auraient en commun le fait d'être reliées au métabolisme du glucose. Mlc serait ainsi un répresseur dont l'activité serait dépendante de la présence de glucose (Plumbridge, 2002). L'opérateur qui permet la fixation de Mlc est positionné de façon à réprimer l'initiation de la transcription des gènes soumis à son contrôle. Les sites opérateurs fixant la protéine Mlc comportent deux éléments symétriques, dont la séquence consensus est très homologue à celle des éléments fixant le répresseur NagC. Cette séquence est (T/A)TTATTT-N9-AAATAA(T/A) (Plumbridge, 2001). Les nucléotides soulignés sont les plus conservés et sont très importants pour la reconnaissance. Il est à noter que les régions de 4 bases notées N4 sont le plus souvent riches en G ou C (Plumbridge, 2001).

Lorsque du glucose est présent dans le milieu de culture, la transcription des gènes cibles de Mlc est augmentée, ce qui signifie que la répression exercée par Mlc est levée.

Le mécanisme de cette levée d'inhibition serait basé sur la séquestration de Mlc au niveau de la membrane par le transporteur spécifique du glucose EIIBC^{Glc} (Fig. 30). En effet, lors du transport du glucose, EIIBC^{Glc} est majoritairement sous forme déphosphorylé, puisque le phosphate est transféré au glucose. Mlc interagit spécifiquement avec EIIBC^{Glc} déphosphorylé qui est situé au niveau de la membrane (Nam *et al.*, 2001; Tanaka *et al.*, 2000; Lee *et al.*, 2000). Dans ce cas, Mlc ne peut plus réprimer la transcription au niveau des gènes du système PTS (Fig. 30 B), en particulier, la transcription du gène *ptsG* est augmentée, ce qui permet la synthèse de nouvelles molécules d'EIIBC^{Glc} qui vont transporter le glucose à l'intérieur de la cellule (Fig. 30 C).

Le domaine B cytoplasmique de la protéine EIIBC^{Glc} est nécessaire à l'interaction avec Mlc (Seitz *et al.*, 2003; Nam *et al.*, 2001) et cette interaction n'a lieu que si ce domaine est lié à la membrane (Seitz *et al.*, 2003; Tanaka *et al.*, 2004). En effet, si le domaine EIIB^{Glc} est produit sous forme soluble, Mlc reste fixée sur les promoteurs même en présence de glucose. Par contre, si le domaine soluble EIIB^{Glc} est fusionné à une autre protéine membranaire comme LacY, la levée de répression est retrouvée pour le gène *ptsG* (Seitz *et al.*, 2003; Tanaka *et al.*, 2004).

La région N-terminale de Mlc serait responsable de l'interaction avec EIIBC^{Glc} (Tanaka *et al.*, 2004) et la région C-terminale de Mlc qui serait impliquée dans l'oligomérisation (tétramérisation) de la protéine, pourrait former une hélice amphipatique (Seitz *et al.*, 2003). Tanaka *et al.* (2004) ont proposé que lorsque la fixation de Mlc sur EIIBC^{Glc} a lieu au niveau de la membrane, l'état tétramérique de la protéine Mlc est modifié. En effet, l'hélice amphipatique formée par la partie C-terminale pourrait s'associer à la membrane, défavorisant ainsi la formation du tétramère actif (Tanaka *et al.*, 2004).

Lorsque le glucose est consommé, la protéine Mlc n'est plus séquestrée par EIIBC^{Glc} et peut réprimer l'expression des protéines du système PTS. En effet, l'affinité de Mlc pour EIIBC^{Glc} sous forme phosphorylée est diminuée. Le phénomène le plus important dans cette diminution d'affinité ne serait pas la modification de la charge de la cystéine catalytique (C421) mais une modification de la structure de cette région lors de la phosphorylation qui empêcherait la fixation de Mlc (Seitz *et al.*, 2003). En plus, Mlc réprime l'expression de son propre gène *mlc*, ce qui permet une autorégulation de sa synthèse en absence de glucose (Nam *et al.*, 2001).

La protéine EIIBC^{Glc} a aussi été décrite comme étant impliquée dans la régulation de la transcription de l'opéron *sdhCDAB* qui code la succinate déshydrogénase, une enzyme du cycle de Krebs (Takeda *et al.*, 1999). La présence de glucose dans le milieu de culture

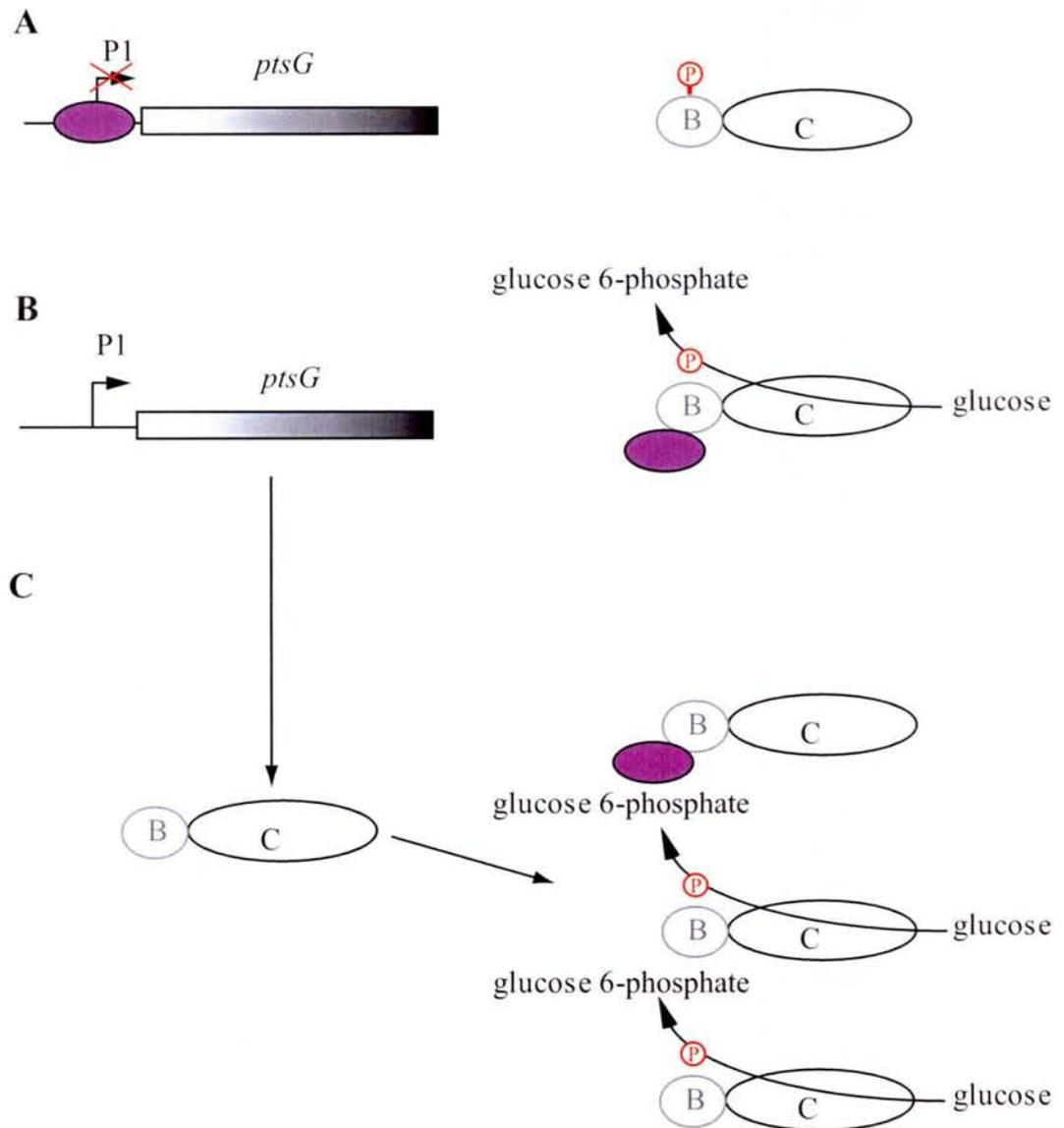


Figure 30: Modèle de la régulation de l'activité des gènes par la protéine Mlc en présence de glucose (d'après Plumbridge, 2002).

A) En absence de glucose, Mlc inhibe l'initiation de la transcription du gène *ptsG* en se fixant au niveau des promoteurs.

B) En présence de glucose, Mlc se fixe sur la protéine EIIBC^{Glc} non phosphorylée, qui est située dans la membrane interne. L'inhibition par Mlc est ainsi levée, ce qui permet la transcription du gène *ptsG* et la production de EIIBC^{Glc}.

C) Les protéines EIIBC^{Glc} nouvellement synthétisées permettent le transport du glucose, alors que Mlc reste fixée au niveau des protéines membranaires EIIBC^{Glc} déphosphorylées.

entraîne une répression de cet opéron et cet effet est dépendant de la présence du transporteur EIIBC^{Glc}. Par contre, le mécanisme permettant cette régulation n'est pas décrit (Takeda *et al.*, 1999).

II.2.2) Les régulations de l'expression des gènes du système PTS.

L'expression des gènes codant les protéines EI et HPr, communes aux différents systèmes PTS et des gènes codant les transporteurs spécifiques des différentes sources de carbone est soumise à un grand nombre de régulations. Ces régulations permettent l'expression des gènes codant les protéines impliquées dans le transport et le métabolisme de ces sources de carbone en fonction des conditions environnementales.

II.2.2.1) Régulation de l'expression des gènes impliqués dans le catabolisme d'autres sucres PTS que le glucose.

Les substrats N-acétyl glucosamine (GlcNAc) et glucosamine (GlcN) peuvent être incorporés dans la cellule par un système PTS spécifique (EIIBC^{Nag}). La protéine EI^{Nag} et les enzymes nécessaires au catabolisme de ces substrats sont codées par un groupe de gènes *nagE-nagBACD* exprimés à partir d'une même région chromosomique par deux promoteurs divergents, *nagE* P1 et *nagB* P1 (Fig. 31). En absence du substrat, l'initiation de la transcription à partir des deux promoteurs est inhibée par le répresseur NagC, qui permet la formation d'une boucle de l'ADN en se fixant au niveau de deux sites appelés boîte E et boîte B (centrées respectivement aux positions -31 et -125 par rapport au site +1 du promoteur *nagE* P1) (Fig. 31) (Plumbridge et Kolb, 1998). La répression est levée en présence d'un inducteur qui pourrait être la N-acétylglucosamine 6-phosphate, produit de la phosphorylation de la N-acétyl glucosamine par son système PTS (Plumbridge, 1990; Plumbridge, 1998a). Cet inducteur permettrait une modification de la structure du répresseur pour qu'il ne se fixe plus sur l'ADN. De plus, un site de fixation pour le complexe CRP-AMPc centré à la position -61,5 par rapport au site d'initiation *nagE* P1 a été mis en évidence (Plumbridge, 1990; Plumbridge, 1991). Ce site est distant du site d'initiation *nagB* P1 de 71,5 pb. La fixation de CRP-AMPc au niveau de cet opérateur permet l'activation des promoteurs divergents *nagE* P1 et *nagB* P1. Le complexe doit alors agir en tant qu'activateur de classe I, certainement en interagissant avec l' α -CTD de l'ARN pol fixée sur chacun des deux promoteurs (Plumbridge, 1990; Plumbridge et Kolb, 1998). Il faut noter que le complexe CRP-AMPc permet de

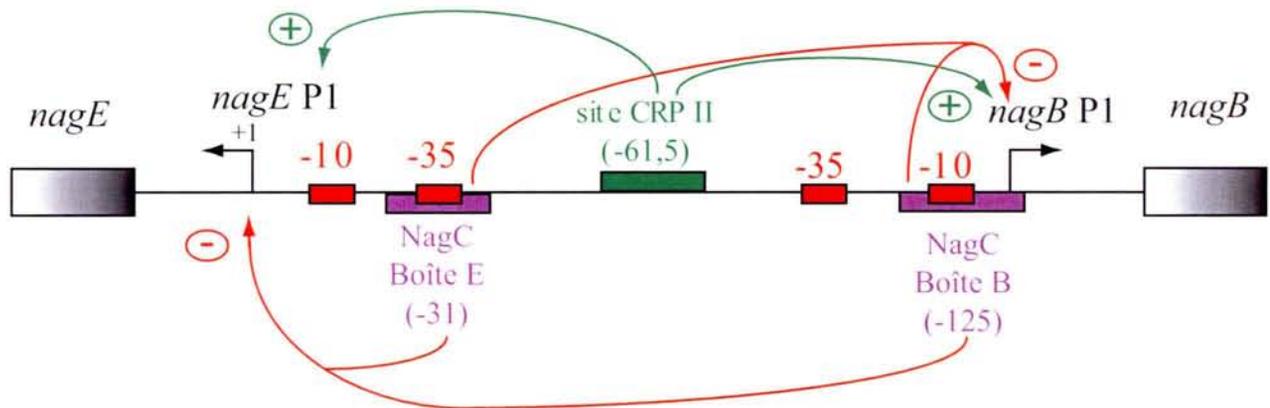


Figure 31 : Représentation schématique de la région promoteur du cluster de gènes *nagE-nagBACD* (d'après Plumbridge et Kolb, 1991 et Plumbridge et Kolb, 1998).

Organisation de la région inter-ORF *nagE-nagB*. Les sites de fixation (boîte E et boîte B) constituant les opérateurs sur lesquels se fixe le répresseur NagC sont indiqués en violet, les boîtes -10 et -35 des promoteurs *nagE* P1 et *nagB* P1 sont représentées en rouge et les sites d'initiation de la transcription par des flèches noires. Le site de fixation du complexe CRP-AMPc est représenté en vert. Les phénomènes d'activation sont indiqués par des flèches vertes (+) et les phénomènes de répression par des flèches rouges (-). La répression de NagC implique la formation d'une boucle de l'ADN entre les deux opérateurs. La fixation du complexe CRP-AMPc peut aussi faciliter la formation de cette boucle.

stabiliser la boucle d'ADN formée entre les répresseurs NagC fixés sur les boîtes B et E (cf. Fig. 23 B) (Plumbridge et Kolb, 1991; Plumbridge et Kolb, 1998). Le gène *nagC* qui est le troisième gène de ce cluster de gènes est aussi transcrit à partir de deux promoteurs propres (*nagC* P1 et P2), situés dans la région codante du gène *nagA*. Ces promoteurs permettent l'expression constitutive de la protéine NagC (Plumbridge, 1996).

La conclusion est que l'expression de cet ensemble de gènes est régulée par la présence du substrat (levée de l'inhibition du répresseur NagC) et par la présence du glucose (diminution du taux d'AMPc et donc du taux du complexe CRP-AMPc).

La plupart des régulations liées aux autres sucres PTS sont basées sur le même principe, elles reposent : 1) sur l'existence d'un répresseur qui empêche l'expression des gènes qui permettent le transport et le métabolisme de ces sources carbonées en absence de ces sources de carbone. 2) sur le complexe CRP-AMPc qui permet l'activation des promoteurs de ces gènes (pour revue : Postma *et al.*, 1993; Postma *et al.*, 1996; Saier *et al.*, 1996).

II.2.2.2) Régulation de la transcription du gène *ptsG* et de l'opéron *pts*.

Le gène *ptsG* est soumis à un grand nombre de régulations. La transcription de ce gène est initiée au niveau de deux promoteurs P1 et P2 (Fig. 32). Le promoteur P1 contribue de façon majoritaire à la transcription du gène *ptsG*, le promoteur P2 ayant une efficacité plus faible (Plumbridge, 1998b). Le complexe CRP-AMPc dont le site de fixation est centré à la position -40,5 en amont du site d'initiation de la transcription est essentiel à l'efficacité du promoteur P1 (Plumbridge, 1998b; Kimata *et al.*, 1997; Kimata *et al.*, 1998). CRP-AMPc serait dans ce cas un activateur de classe II et pourrait faciliter la formation des complexes ouverts au niveau du promoteur *ptsG* P1 (Shin *et al.*, 2003). Un deuxième site de fixation pour CRP-AMPc, centré à la position -95, a été détecté en amont du promoteur *ptsG* P1. Le rôle de ce site n'a pas encore été élucidé (Kimata *et al.*, 1998). Il est à noter que la région -35 du promoteur P1 ne présente d'homologie avec la séquence consensus des boîtes -35. L'absence de boîte -35 serait compensé par la fixation du complexe CRP-AMPc (Plumbridge, 1998b).

En présence de glucose dans le milieu de culture, l'initiation de la transcription au niveau du promoteur P1 est augmentée. Cette augmentation est liée à la levée de l'inhibition exercée par la protéine Mlc, dont le site de fixation est centré à la position -6 par rapport au site d'initiation P1. Le mécanisme de cette levée d'inhibition a été décrit dans la partie précédente (Fig. 30). De la même manière, *ptsG* P2 serait régulé négativement par Mlc, dont le site de

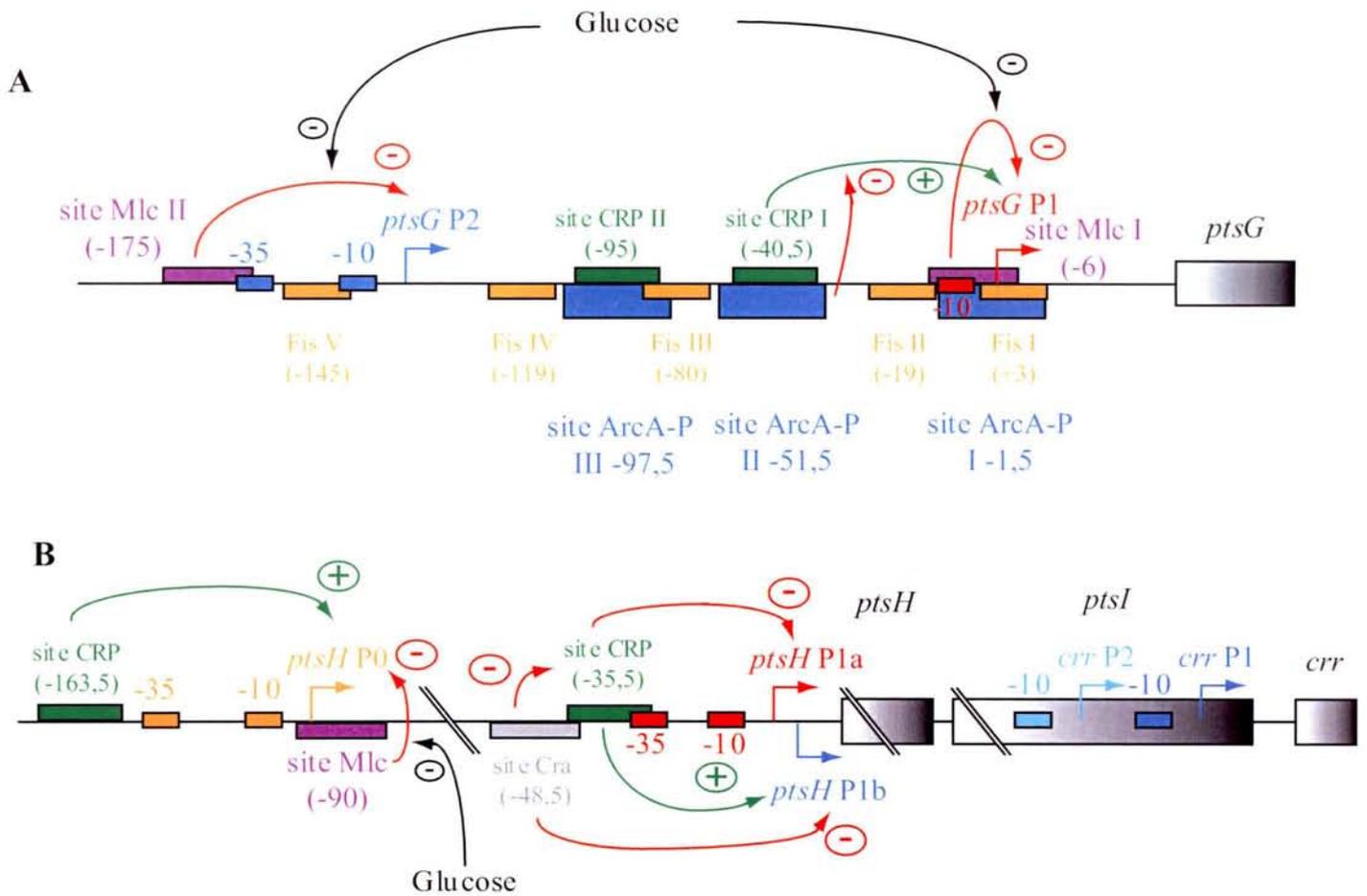


Figure 32 : Représentation schématique des régions promotrices du gène *ptsG* et de l'opéron *pts*.

A) Régulation de l'initiation de la transcription du gène *ptsG* (d'après Plumbridge, 1998 ; Shin *et al.*, 2003 ; Jeong *et al.*, 2004). Le promoteur majoritaire *ptsG* P1 et sa boîte -10 sont indiqués en rouge et le promoteur P2 en bleu. Les sites de fixation du complexe CRP-AMPC sont représentés en vert, les sites de fixation de Fis en orange et de ArcA-P en bleu. Les phénomènes d'activation sont indiqués par des flèches vertes (+) et les phénomènes de répression par des flèches rouges (-). L'effet du glucose qui permet la levée de l'inhibition de Mlc est indiqué par un flèche noire (-). Les effets de Fis ne sont pas indiqués sur la figure mais décrits dans le texte. Les sites de fixation de ArcA-P chevauchent le site d'initiation de la transcription P1 et les sites de fixation du complexe CRP-AMPC.

B) Régulation de l'initiation de la transcription de l'opéron *pts* (d'après De Reuse et Danchin, 1988 ; Ryu *et al.*, 1995 ; Plumbridge, 1999). Les promoteurs sont indiqués par des couleurs différentes. Les sites de fixation pour les facteurs de transcription sont représentés de la même façon qu'en A). Le site de fixation pour la protéine Cra est représenté en gris. Les phénomènes d'activation sont indiqués par des flèches vertes (+) et les phénomènes de répression par des flèches rouges (-). L'effet du glucose qui permet la levée de l'inhibition de Mlc est indiqué par un flèche noire (-).

fixation est centré 35 pb en amont du site P2 (Shin *et al.*, 2003). De plus, la transcription du gène *ptsG* est régulée par les protéines Fis (Shin *et al.*, 2003) et ArcA (Jeong *et al.*, 2004). Ces régulations lient donc l'expression du gène *ptsG* à la phase de croissance par l'intermédiaire de Fis et à la disponibilité en oxygène dans le milieu par ArcA.

En absence de glucose, la protéine Fis pourrait faciliter la fixation de Mlc sur ses opérateurs en favorisant la formation de complexes nucléoprotéiques CRP-AMPC, Fis et Mlc et aurait ainsi un effet négatif sur l'initiation de la transcription au niveau des deux promoteurs (Shin *et al.*, 2003). Par contre, en présence de glucose, l'inhibition exercée par Mlc est levée et Fis pourrait faciliter la fixation du complexe CRP-AMPC. Dans ce cas, l'effet de Fis serait une activation de la transcription à partir du promoteur P1. Le renforcement par Fis de la répression par Mlc et de l'activation par CRP-AMPC permettrait de lier la production de EIIBC^{Glc} à la phase de croissance. De plus, comme le taux de CRP-AMPC diminue en présence de glucose, Fis pourrait aider à augmenter le taux d'occupation du site CRP-AMPC pour augmenter l'initiation de la transcription au niveau du promoteur *ptsG* P1.

La protéine ArcA sous sa forme phosphorylée (ArcA-P) se fixe sur cette région. Pour mettre en évidence cette fixation, Jeong *et al.* (2004) ont utilisé une technique de « ligand fishing ». Cette technique consiste à fixer sur une colonne de streptavidine, des fragments d'ADN amplifiés par PCR avec des oligonucléotides biotinylés, ceci à partir de la séquence du promoteur. Ensuite, un extrait cytoplasmique d'*E. coli* est déposé sur la colonne et les protéines qui se fixent sur l'ADN du promoteur sont retenues et identifiées par spectrométrie de masse. La protéine ArcA fait partie du système à deux composants ArcA/ArcB qui permet l'adaptation d'*E. coli* à des changements de conditions respiratoires. En effet, en condition de croissance anaérobie, la protéine ArcB s'autophosphoryle et transmet son phosphate à la protéine ArcA qui est le régulateur de réponse. Cette dernière acquiert la capacité à se fixer à l'ADN et à réguler négativement l'expression de gènes et d'opérons impliqués dans le métabolisme respiratoire (enzymes de transport d'électrons, enzymes du cycle de Krebs) et à réguler positivement l'expression des gènes impliqués dans le métabolisme fermentaire (pyruvate formate lyase) (pour revue : Sawers, 1999; Georgellis *et al.*, 2001). En condition de croissance aérobie, les quinones du système de transport des électrons, qui se trouvent majoritairement sous forme oxydée du fait de la présence d'un accepteur d'électrons, peuvent inhiber l'autophosphorylation de ArcB et ainsi modifier le profil de régulation de l'expression des gènes contrôlés par ArcA (Georgellis *et al.*, 2001).

Au niveau du promoteur *ptsG* P1, les sites ArcA-P sont centrés aux positions +1,5 (ArcA-P I), -51,5 (ArcA-P II) et -97,5 (ArcA-P III) (Fig. 32). La fixation de ArcA-P au niveau de ce

promoteur inhibe l'initiation de la transcription (Jeong *et al.*, 2004). L'élément ArcA-P I chevauche le site d'initiation de la transcription, alors que les sites II et III chevauchent les sites de fixation des complexes CRP-AMPc (Fig. 32). En effet, ArcA-P et CRP-AMPc sont en compétition pour la fixation sur l'ADN du promoteur. Les auteurs concluent qu'en absence d'oxygène, ArcA-P inhibe la transcription du gène *ptsG* pour limiter l'entrée du glucose du fait que la phosphorylation oxydative est limitée. Ce mécanisme pourrait ainsi limiter la production d'acides par la glycolyse (Jeong *et al.*, 2004).

Il est à noter que la stabilité des ARNm *ptsG* est fortement régulée par le métabolisme du glucose. En effet, lorsque la glycolyse est bloquée et que le glucose 6-phosphate et/ou le fructose 6-phosphate s'accumulent (par exemple par mutation des gènes *pgi*, *pfkA* ou *fbaA*), la dégradation des ARNm *ptsG* par l'activité de la RNase E est beaucoup plus rapide (Kimata *et al.*, 2001; Morita *et al.*, 2003). Il existe donc un lien entre le flux métabolique au niveau de la glycolyse et la dégradation des ARNm codant le système de perméation spécifique du glucose.

L'activation par le complexe CRP-AMPc du gène *ptsG* a aussi un effet indirect sur l'exclusion de l'inducteur au niveau du gène *lac*. En présence de glucose et de lactose, le glucose est consommé préférentiellement, ce qui diminue le taux de phosphorylation de EIIA^{Glc} par l'intermédiaire de EIIBC^{Glc}. EIIA^{Glc} favorise l'exclusion de l'inducteur. Le répresseur LacI limite l'initiation de la transcription au niveau du promoteur *lac*. Par contre, CRP-AMPc est essentiel à la transcription de *ptsG* qui permet la production de EIIBC^{Glc} induisant la déphosphorylation de EIIA^{Glc}. Le complexe CRP-AMPc serait donc important pour l'exclusion de l'inducteur par son effet sur la transcription de *ptsG* (Kimata *et al.*, 1997).

La transcription de l'opéron *pts* (gènes *ptsH*, *ptsI* et *crr*) est initiée à partir de trois promoteurs *ptsH* P0, *ptsH* P1a et *ptsH* P1b pour les gènes *ptsH* et *ptsI*. Le gène *crr* est transcrit à partir de ces promoteurs mais aussi à partir de deux promoteurs spécifiques *crr* P1 et *crr* P2, localisés dans la région codante de *ptsI* (Fig. 32) (De Reuse et Danchin, 1988; De Reuse *et al.*, 1992; Fox *et al.*, 1992; Ryu et Garges, 1994). Comme dans le cas du gène *ptsG*, l'initiation de la transcription à partir du promoteur *ptsH* P0 est régulée positivement par le complexe CRP-AMPc et négativement par la protéine Mlc (De Reuse *et al.*, 1992; Plumbridge, 1999). Il est à noter que d'après sa séquence, ce promoteur posséderait une boîte -10 étendue. En présence de glucose dans le milieu de culture, la répression par Mlc est levée (Fig. 30) (Plumbridge, 1999). Le site de fixation de la protéine Mlc chevauche le site d'initiation de la transcription. De plus, l'initiation de la transcription au promoteur *ptsH* P0 est strictement dépendante du

complexe CRP-AMPc qui se fixe sur un site centré 63,5 pb en amont du site +1 de l'initiation (De Reuse et Danchin, 1988; Ryu et Garges, 1994; Plumbridge, 1999). Cette position du site de fixation de CRP-AMPc est caractéristique d'un activateur de classe I.

La régulation au niveau des promoteurs *ptsH* P1 est plus complexe que celle de *ptsG* P1 et fait intervenir en plus de CRP-AMPc, la protéine Cra. Cette protéine appelée aussi FruR, régule l'expression de gènes codant des enzymes impliquées dans la glycolyse, la néoglucogenèse, le cycle de Krebs et le shunt du glyoxylate (Saier et Ramseier, 1996). La fixation de la protéine Cra à son opérateur est inhibée par le fructose 1-phosphate, qui est le produit du transport du fructose par le système PTS spécifique de ce composé. La fixation de cette protéine est aussi inhibée par le fructose 1,6-diphosphate, qui est un intermédiaire de la glycolyse (Ramseier *et al.*, 1995; Ramseier *et al.*, 1993). L'effet régulateur de la protéine Cra est donc lié au transport et au métabolisme des sources de carbone dans la cellule. Cette protéine est à la fois activateur et répresseur suivant la localisation de son site de fixation par rapport au promoteur. Le mécanisme de l'activation de la transcription ne semble pas basé sur l'interaction de la protéine Cra avec l'ARN pol mais sur sa capacité à faciliter l'interaction de l' α -CTD avec l'ADN en créant une courbure de l'ADN (Nègre *et al.*, 1998; Prost *et al.*, 1999). Ce facteur de transcription influence les flux métaboliques au niveau de la glycolyse et de la néoglucogenèse. En effet, Cra limite l'expression des gènes de la glycolyse, alors qu'elle active l'expression des gènes de la néoglucogenèse (pour revue : Saier et Ramseier, 1996).

Dans le cas du promoteur *ptsH* P1, la fixation du complexe CRP-AMPc en amont des deux promoteurs P1a et P1b dont les sites +1 sont séparés par 7 pb, régule de manière différente l'utilisation de chacun de ces promoteurs. Le complexe réprime l'utilisation du promoteur P1a (site de fixation situé à -35,5 du site d'initiation) et active celle du promoteur P1b (site de fixation situé à -42,5 du site d'initiation) (Ryu et Garges, 1994; De Reuse *et al.*, 1992). La fixation de la protéine Cra sur son opérateur limite la fixation du complexe CRP-AMPc au niveau de son opérateur (Ryu, 1995) favorisant ainsi l'initiation au site P1a. Ces différentes régulations semblent maintenir un taux d'initiation de la transcription constant dans un grand nombre de conditions de croissance (Ryu et Garges, 1994; De Reuse *et al.*, 1992).

Les promoteurs *crr* P1 et P2 font partie de la famille des promoteurs étendus et ne seraient pas régulés. Ils permettent une production constitutive de EIIA^{Glc}.

En conclusion, la transcription de l'opéron *pts* est initiée à partir de deux types de promoteurs : 1) des promoteurs non régulés, qui permettent une expression de base des protéines EI, HPr et EIIA^{Glc} pour permettre le transport des sources de carbone PTS, mais aussi une expression de base de la protéine EIIA^{Glc} pour qu'elle joue son rôle de régulateur.

2) un promoteur régulé qui permet l'augmentation de la production des trois protéines lorsque du glucose est présent dans le milieu de culture.

II.2.2.3) Régulation de la transcription des gènes *crp* et *cya*.

La transcription du gène *cya* qui code l'adénylate cyclase est initiée à partir de trois promoteurs P1, P1' et P2 (Fig. 33 A).

L'initiation de la transcription à partir du promoteur majoritaire P2 est inhibée par le complexe CRP-AMPc dont le site de fixation chevauche le site +1 (Aiba, 1985). En présence de glucose, le taux de CRP-AMPc diminue, ce qui lève partiellement l'inhibition. Le taux d'adénylate cyclase est donc augmenté mais son activité serait faible du fait que, dans ces conditions physiologiques, la protéine EIIA^{Glc} est majoritairement déphosphorylée. Cette double régulation permettrait une adaptation rapide au changement de nutriments après la consommation du glucose. En effet, un taux élevé d'adénylate cyclase permet la synthèse d'AMPc dès lors que EIIA^{Glc}-P est présent, ce qui entraîne une activation de la transcription des gènes impliqués dans le métabolisme des sources de carbone alternatives (Saier et Ramseier, 1996).

Le gène *crp* qui code la protéine CRP est fortement régulé. La transcription est initiée à partir de deux promoteurs P1 et P2. Le promoteur P2 est non canonique et serait régulé négativement par fixation de la protéine Fis (fixée au site I), mais aussi par le taux de surenroulement de l'ADN (Gonzalez-Gil *et al.*, 1998).

Deux sites de fixation pour le complexe CRP-AMPc sont présents au niveau de cette région promotrice (Fig. 33 B) (Hanamura et Aiba, 1992; Aiba, 1983). Le premier site (II) est centré à la position -61,5 par rapport au site d'initiation du promoteur P1, tandis que le deuxième (I) est situé en aval du promoteur à la position +42. La fixation de CRP-AMPc sur le site I semble réprimer l'initiation de la transcription à partir du promoteur P1 (Aiba, 1983; Hanamura et Aiba, 1991), alors que la fixation au niveau du site II permet une activation de la transcription à partir de ce promoteur (Hanamura et Aiba, 1992; Ishizuka *et al.*, 1994; Ishizuka *et al.*, 1993). Il est à noter que le site I a une affinité plus forte pour CRP-AMPc que le site II (Pyles et Lee, 1996). Ishizuka *et al.* (1994) ont proposé un modèle pour expliquer l'effet de la présence de glucose dans le milieu de culture. Le transport du glucose permet la déphosphorylation de EIIA^{Glc}, ce qui entraîne une diminution du taux d'AMPc intracellulaire. La concentration de CRP-AMPc diminue et, dans ce cas, seul le site I de plus forte affinité

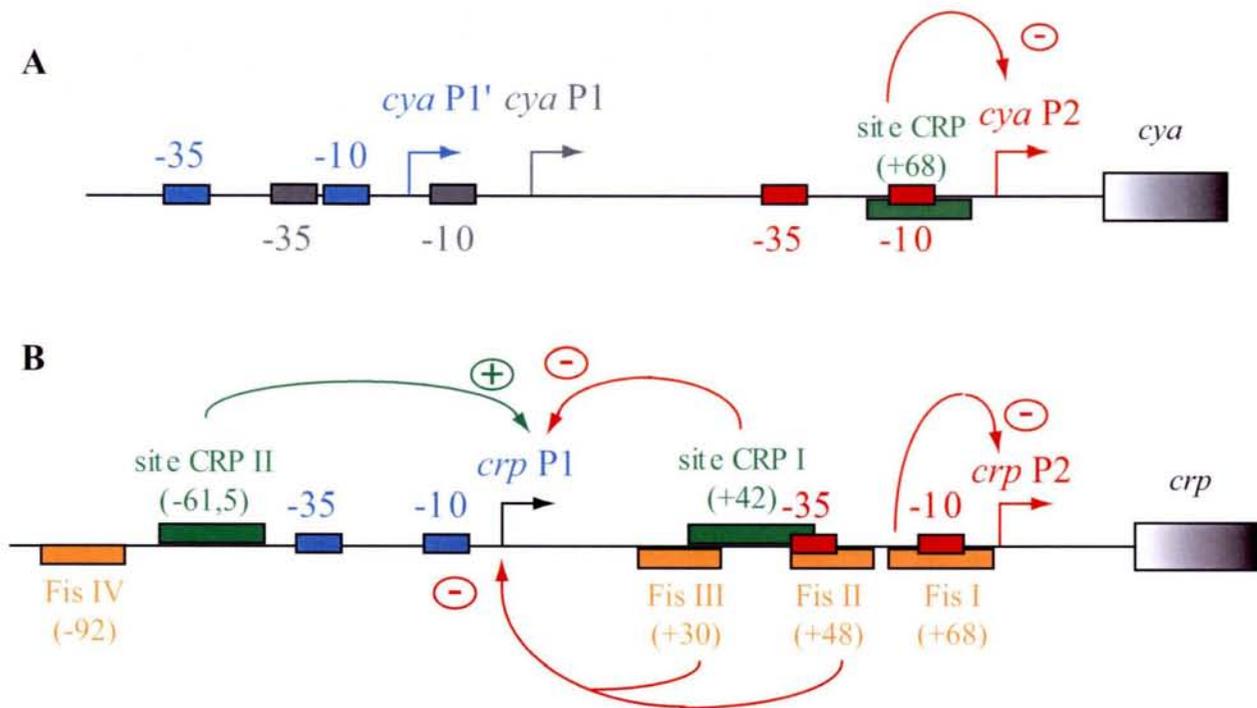


Figure 33 : Représentation schématique des régions promotrices des gènes *cya* et *crp* (d'après Aiba, 1985; Hanamura et Aiba, 1991; Ishizuka *et al.*, 1994 et Gonzalez-Gil *et al.*, 1998).

A) Régulation de l'initiation de la transcription du gène *cya*. Le promoteur majoritaire *cya* P2, ainsi que les boîtes -10 et -35, sont indiqués en rouge. Les promoteurs minoritaires P1 et P1' sont indiqués respectivement, en gris et bleu. Le site de fixation du complexe CRP-AMPc est représenté en vert. La fixation de CRP-AMPc au niveau de cet opérateur entraîne une inhibition de l'initiation de la transcription indiquée par une flèche rouge (-).

B) Régulation de l'initiation de la transcription du gène *crp*. Les promoteurs P1 et P2 sont représentés en bleu et rouge. Les sites de fixation des complexes CRP-AMPc sont représentés en vert et les sites de fixation de la protéine Fis en orange. Les phénomènes d'activation sont indiqués par des flèches vertes (+) et les phénomènes de répression par des flèches rouges (-).

serait occupé, ce qui entraînerait majoritairement une inhibition de l'initiation de la transcription à partir du promoteur P1.

La protéine Fis possède trois sites de fixation au niveau de cette région promotrice (Gonzalez-Gil *et al.*, 1998). Ces sites chevauchant le site de fixation de CRP-AMPc, leur occupation par Fis entraînerait une inhibition de l'initiation de la transcription (Gonzalez-Gil *et al.*, 1998). Le site ayant l'affinité la plus élevée pour Fis est le site I qui inhiberait la transcription à partir des deux promoteurs. La régulation de la transcription du gène *crp* dépend donc de trois paramètres : les concentrations du complexe CRP-AMPc, de la protéine Fis et le taux de surenroulement.

Comme nous venons de le décrire, les systèmes PTS de transport des sources de carbone sont centraux au niveau du métabolisme d'*E. coli*. Les régulations qu'exercent les protéines qui composent ces systèmes sont essentielles pour l'établissement de la hiérarchie de consommation des composés carbonés, mais aussi pour limiter le transport et le métabolisme des sources de carbone en fonction des capacités métaboliques de la cellule. Ces régulations se produisent aussi bien au niveau protéique grâce à des interactions entre protéines (par exemple, EIIA^{Glc}-perméases) dont l'activité varie en fonction des conditions environnementales, qu'au niveau génétique grâce à des facteurs de transcription dont l'activité et/ou la concentration varient aussi en fonction des conditions du milieu de culture, et qui interagissent avec l'ADN et régulent l'initiation de la transcription. La répression de l'utilisation des sources de carbone alternatives basée sur un répresseur inductible par la source de carbone et sur un facteur transcriptionnel dont la concentration varie en fonction de la source de carbone et du métabolisme permet de limiter le nombre de gènes transcrits et assure une utilisation économique des nutriments présents dans l'environnement de la bactérie. Ces systèmes permettent donc une adaptation optimale d'*E. coli* assurant la survie dans des milieux contenant différentes sources de carbone dont la concentration peut fortement varier.

III.) Les systèmes à deux composants.

Les bactéries sont confrontées à la présence de différentes molécules qui se trouvent dans leur environnement mais qui sont aussi produites par leur métabolisme. Les cellules possèdent des systèmes de détection qui induisent une réponse d'adaptation appropriée à la présence de ces molécules. Ces systèmes permettent de transformer une reconnaissance moléculaire en un signal qui régule l'expression de gènes et d'autres processus cellulaires afin d'adapter la cellule à son environnement. Les systèmes à deux composants sont aussi impliqués dans l'expression de protéines essentielles pour la pathogénicité des bactéries à gram⁻. Dans le cas des systèmes à deux composants, le signal est une phosphorylation d'un des deux acides aminés suivants, l'histidine et l'acide aspartique portés par des protéines. Ces systèmes qui ont déjà été décrits auparavant ont été appelés phospho-relais (pour revue : Egger *et al.*, 1997; Hoch, 2000; Stock *et al.*, 2000; West et Stock, 2001). Ces phospho-relais sont composés d'un senseur membranaire appelé histidine kinase (HK), qui possède un domaine kinase permettant l'autophosphorylation et un domaine de reconnaissance pour un signal spécifique. Lors de l'activation du senseur par le signal, le domaine kinase peut s'autophosphoryler au niveau d'une histidine. Le phosphate est ensuite transféré sur l'acide aspartique d'un régulateur de réponse à des signaux externes (RR). Ce régulateur est constitué de deux domaines, le domaine receveur (ou régulateur) et le domaine effecteur. Le rôle du domaine receveur non phosphorylé est d'inhiber l'activité du domaine effecteur. Lors de la phosphorylation par le senseur, cette inhibition est levée et le domaine effecteur devient actif. Cette fonction est en général une fixation sur l'ADN, ce qui permet ainsi une régulation de l'initiation de la transcription.

Après une présentation des différents systèmes à deux composants chez les bactéries, je décrirai plus en détails le système EnvZ/OmpR, qui répond à l'osmolarité du milieu et le système CheA/CheY qui répond à la présence de nutriments et régule la mobilité de la bactérie.

III.1) L'organisation génomique des systèmes à deux composants.

Ces systèmes sont retrouvés dans les organismes des trois règnes : bactéries, Archaea et eucaryotes. Par contre, ils sont rares chez les eucaryotes et très fréquents chez les bactéries. Chez les eucaryotes, les systèmes de transfert d'information sont plutôt basés sur la

phosphorylation des acides aminés sérine, thréonine et tyrosine (pour revue : Stock *et al.*, 2000; West et Stock, 2001). Parmi les eucaryotes, des systèmes à deux composants ont été identifiés seulement chez les plantes, les protozoaires, les levures et les champignons (Wolanin *et al.*, 2002). Les génomes de l'homme et de la drosophile ne codent aucun système à deux composants reposant sur le transfert His-Asp (West et Stock, 2001).

III.1.1) Chez les bactéries.

Le nombre de protéines identifiées comme appartenant à des systèmes à deux composants varie très fortement suivant les bactéries. Ce nombre a été estimé à partir de l'analyse des séquences des génomes bactériens qui ont été entièrement séquencés (pour revue : Stock *et al.*, 2000; West et Stock, 2001). Ces études montrent que certains organismes ne possèdent pas de protéines de systèmes à deux composants comme *Mycoplasma genitalium* et que d'autres peuvent posséder jusqu'à 80 protéines comme la cyanobactérie *Synechocystis* sp. (Mizuno *et al.*, 1996; Mizuno, 1998).

III.1.2) Chez *E. coli*.

Une étude informatique a permis de détecter chez *E. coli* les protéines qui possèdent une homologie avec des protéines connues de systèmes à deux composants (Mizuno, 1997). Dans le génome d'*E. coli*, il existe 62 ORF qui feraient partie de systèmes à deux composants. Parmi ces protéines, 32 seraient des régulateurs de réponse à des signaux externes, 25 impliqueraient des histidine kinase et 5 seraient des histidine kinase hybrides (Mizuno, 1997). Parmi ces protéines, certaines avaient déjà été identifiées expérimentalement comme faisant partie de systèmes à deux composants.

Une prédiction de l'ensemble des différents systèmes à deux composants chez *E. coli* est accessible sur le site <http://www.genome.jp/kegg/ortholog/tab02020.html>. Ce site regroupe les 62 ORF d'*E. coli* réparties en familles de régulateurs de réponse à des signaux externes. Les histidine kinases permettant la phosphorylation de ces RR sont indiquées ainsi que les conditions d'activation de ces systèmes à deux composants. Ce site donne la liste des systèmes à deux composants présents chez d'autres organismes.

Je vais maintenant détailler les structures et les fonctions de ces différents systèmes chez *E. coli*.

III.2) Structure et fonction des systèmes à deux composants d' *E. coli*.

III.2.1) Le composant histidine kinase.

La séquence des domaines senseurs varie en fonction de l'HK, ce qui implique une activation par des stimuli différents. Le mécanisme d'activation, ainsi que les signaux qui permettent l'activation sont peu connus (pour revue : Stock *et al.*, 2000; West et Stock, 2001).

Le domaine kinase est composé d'un domaine d'autophosphorylation qui permet aussi la dimérisation de l'HK et d'un domaine qui lie l'ATP et possède l'activité catalytique (Fig. 34). Le domaine senseur permet de moduler l'activité du domaine kinase dont l'activité d'autophosphorylation ATP dépendante s'exerce sur un résidu histidine conservé du domaine de dimérisation (représenté par un H ; Fig. 34 et 35). L'autophosphorylation est réalisée entre les deux dimères c'est-à-dire que le domaine catalytique d'un monomère phosphoryle l'histidine conservée de l'autre monomère (Swanson *et al.*, 1993; Ninfa *et al.*, 1993; Stock *et al.*, 2000). Certaines HK possèdent une activité phosphatase, ce qui leur permet de déphosphoryler le régulateur de réponse (Keener et Kustu, 1988; Aiba *et al.*, 1989). Cette activité phosphatase permet ainsi un retour rapide à un taux de phosphorylation faible pour le régulateur de réponse et donc un arrêt de la transmission du signal.

Chez *E. coli*, il existe différentes familles d'histidine kinases (Fig. 34 et 35).

Dans les systèmes à deux composants typiques (Fig. 34 A et 35 A), l'histidine kinase est insérée dans la membrane interne et détecte un stimulus périplasmique. Ce type d'organisation est retrouvé pour l'HK EnvZ et pour un grand nombre d'autres protéines homologues (Fig. 35 A) (Mizuno, 1997) (pour revue : Stock *et al.*, 2000; West et Stock, 2001). Ces HK sont constituées d'un domaine senseur, qui se trouve dans le périplasme et du domaine kinase qui est cytoplasmique.

Il est à noter que toutes les histidine kinases ne sont pas membranaires. En effet, l'HK CheA, qui fait partie du système CheA/CheY lié au chimiotactisme bactérien, est cytoplasmique. Elle peut s'associer à des récepteurs membranaires qui permettent son activation. L'histidine kinase CheA possède une structure différente des autres HK (Fig. 36). En effet, l'histidine qui est phosphorylée se situe au niveau de l'extrémité N-terminale de la protéine, au niveau d'un domaine HPt (Histidine phosphotransférase), appelé P1, qui contient une histidine phosphorylable par le domaine kinase P4 (Fig. 36). Le site de dimérisation se situe au niveau du domaine P3, alors que le site de liaison aux récepteurs membranaires se situe au niveau du domaine P5. De plus, une propriété de CheA est qu'elle peut phosphoryler deux cibles

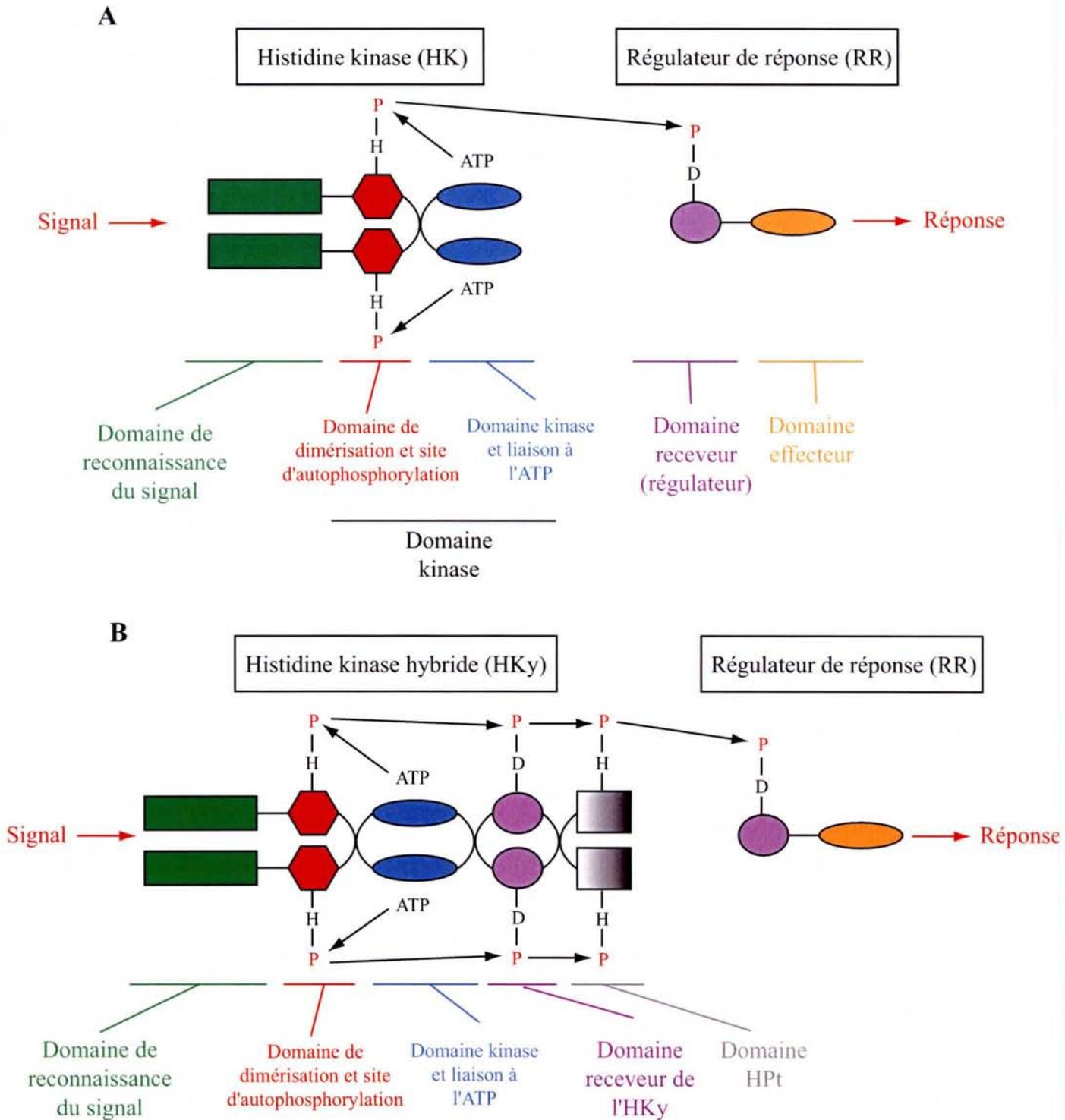


Figure 34 : Organisation des systèmes à deux composants et transfert du signal (d'après Hoch, 2000 ; Stock *et al.*, 2000 ; West et Stock, 2001).

Les domaines sont représentés avec des couleurs différentes : le domaine de reconnaissance du signal en vert, le domaine kinase en rouge (il contient l'histidine phosphorylée et le site de dimérisation) et le domaine de fixation de l'ATP qui possède l'activité kinase en bleu, les domaines régulateurs sont en violet et les domaines effecteurs en orange. Les phosphates sont représentés par un P rouge et les lettres H et D représentent les acides aminés histidine et acide aspartique, qui sont phosphorylés. Le domaine HPT (Histidine phosphotransférase) est représenté par un dégradé de gris.

A) Le système EnvZ/ OmpR.

B) Le système ArcB/ArcA. Le phosphate est transféré séquentiellement à partir du domaine kinase vers un domaine régulateur, puis à un autre domaine de phosphotransfert, qui font partie de la même protéine. Le phosphate est ensuite transféré à la protéine régulateur de réponse.

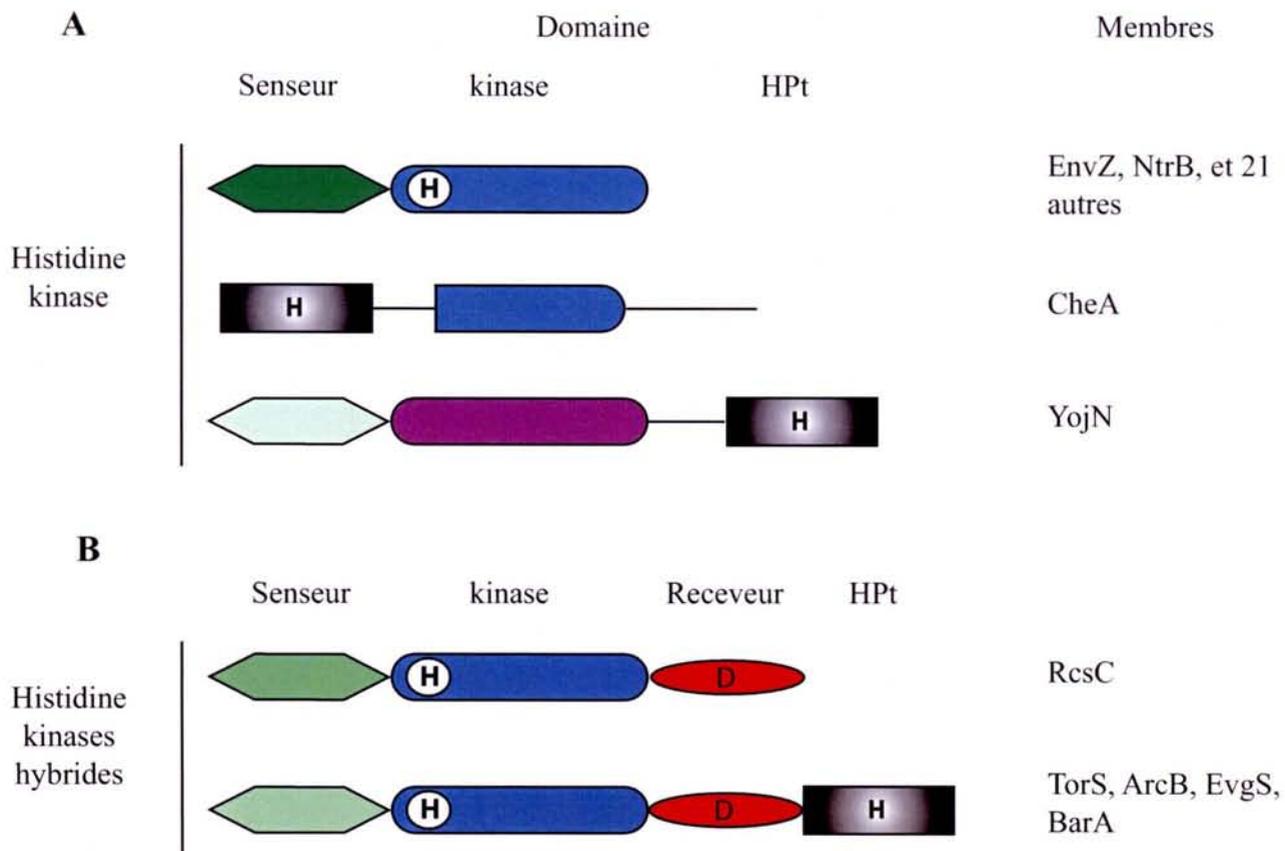


Figure 35 : Organisation des histidine kinases d'*E. coli* (d'après Mizuno, 1997).
 Les domaines senseurs, kinases, receveurs et HPt sont représentés avec des couleurs différentes pour chaque famille d'histidine kinase. Les membres de ces familles sont indiqués sur la droite.

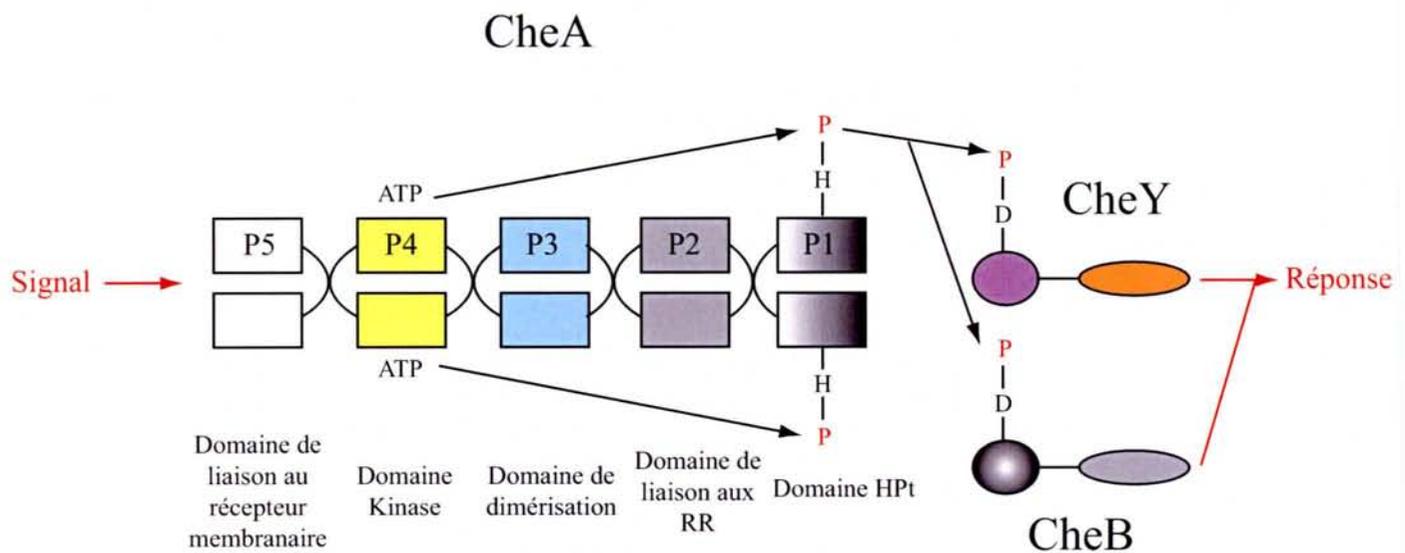


Figure 36 : Organisation du système à deux composants CheA/CheY (d'après Hoch, 2000 ; Stock *et al.*, 2000 ; West et Stock, 2001).

Les domaines P1 à P5 qui portent les différentes activités sont séparés et représentés avec des couleurs différentes. Ce type d'HK peut transférer son phosphate à deux cibles RR différentes (CheY et CheB). Les domaines régulateurs des RR sont représentés en violet et gris foncé et les domaines effecteurs en orange et gris clair. Les phosphates sont représentés par un P rouge et les lettres H et D représentent les acides aminés histidine et acide aspartique qui sont phosphorylés. Le domaine HPt (Histidine phosphotransférase) est représenté par un dégradé de gris.

différentes, alors qu'en général, les histidine kinases ont une spécificité stricte et ne phosphorylent que leur régulateur de réponse (Fig. 36). Cette reconnaissance de deux RR se fait par l'intermédiaire du domaine P2 (pour revue : Manson *et al.*, 1998; Stock *et al.*, 2000; West et Stock, 2001).

Dans le transfert du phosphate au régulateur de réponse RcsB, il existe un intermédiaire, la protéine YojN, qui est phosphorylée au niveau de son domaine HPt par RcsC et qui pourrait transmettre le phosphate au domaine receveur du régulateur de réponse (Chen *et al.*, 2001; Takeda *et al.*, 2001) (Fig. 35).

Il existe aussi une famille d'histidine kinases appelées hybrides (HKy), qui sont composées de plusieurs domaines (Ishige *et al.*, 1994) (Fig. 34 B et 35 B). Une HKy typique de cette famille chez *E. coli* est ArcB qui est sensible aux changements de conditions respiratoires de la bactérie (Fig. 34 B et 35 B) (pour revue : Stock *et al.*, 2000). Cette protéine possède un domaine senseur membranaire, un domaine kinase, un domaine ayant une homologie avec un domaine receveur (avec un acide aspartique essentiel) et un domaine HPt qui contient une histidine permettant le transfert du phosphate au régulateur de réponse, ArcA. Après activation de l'activité catalytique kinase par le domaine senseur, ArcB s'autophosphoryle sur une histidine située au niveau du domaine kinase. Le phosphate est ensuite transféré sur l'acide aspartique du domaine receveur de l'HKy. Ce phosphate peut ensuite être transféré sur l'histidine du domaine HPt de l'HKy puis vers le régulateur de réponse (Fig. 34 B). Chaque étape de transfert de phosphate entre les domaines de l'HKy peut être régulée, ce qui permet une intégration de plusieurs signaux de régulation en une seule réponse liée à la phosphorylation du régulateur de réponse ArcA (pour revue : Stock *et al.*, 2000). Dans cette même famille d'HK hybrides chez *E. coli*, la protéine RcsC possède un domaine receveur supplémentaire qui porte un acide aspartique essentiel. Le phosphate peut être transféré à la protéine YojN qui pourra le transmettre au régulateur de réponse RcsB. D'autres exemples montrent que les différents domaines qui transmettent le phosphate peuvent être solubles. C'est le cas du système de sporulation chez *Bacillus subtilis*.

III.2.2) Régulateurs de réponse.

Les régulateurs de réponse sont les effecteurs qui permettent une réponse appropriée aux conditions environnementales. Comme mentionné ci-dessus, ils sont composés de deux domaines : un domaine receveur (aussi appelé régulateur) et un domaine effecteur (Fig. 34 et 37).

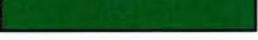
Famille	Domaine		Membres
	Receveur (régulateur)	Effecteur	
CheY			CheY
OmpR			PhoB, CusR, KdpE, TorR, PhoP, RstA, YedW, BaeR, QseB, ArcA, OmpR, CpxR, BasR, CreB
NarL			FimZ, NarL, UvrY, RcsB, EvgA, NarP, UhpA
NtrC			AtoC, YfhA, NtrC, HydG
Autres			CitB, YpdB, DcuR, YehT
CheB			CheB, RssB

Figure 37 : Organisation des régulateurs de réponse d'*E. coli* (d'après Mizuno, 1997).

Les domaines receveurs et accepteurs sont représentés avec des couleurs différentes pour chaque famille de régulateur de réponse. Les membres de ces familles sont indiqués sur la droite.

La plupart des RR sont des facteurs de transcription (pour revue : Kenney, 2002). En effet, les domaines effecteurs possèdent, pour la plupart, la capacité de se fixer à l'ADN (25 régulateurs sur 32 RR chez *E. coli* sont prédits se lier à l'ADN) (pour revue : Stock *et al.*, 2000). Les séquences ADN reconnues par chacun de ces facteurs sont différentes.

Il existe plusieurs familles de RR (Fig. 37). Ces familles diffèrent par la séquence des domaines effecteurs. La famille la plus représentée est la famille OmpR qui contient 14 membres (Kenney, 2002), puis la famille NarL avec 7 membres et la famille NtrC avec 4 membres. Le RR CheY n'est constitué que d'un domaine receveur. C'est la fixation de ce domaine au niveau du moteur du flagelle, qui permet de réguler la direction de sa rotation.

D'autres RR sont constitués d'un domaine effecteur qui possède une activité enzymatique, comme CheB qui possède une activité méthyle transférase (Simms *et al.*, 1985). Les autres RR ont été très peu étudiés (Fig. 37).

Le mécanisme d'activation du domaine effecteur par le domaine receveur est basé sur la phosphorylation (Robinson *et al.*, 2000). Les RR catalysent le transfert du phosphate à partir de l'histidine phosphorylée de l'HK jusqu'à l'acide aspartique essentiel situé au niveau du domaine receveur. Cette phosphorylation pourrait entraîner un changement de conformation du RR (Robinson *et al.*, 2000; Stock *et al.*, 2000; Kenney, 2002).

En effet, la structure d'un RR pourrait osciller entre une forme active et une forme inactive. La phosphorylation pourrait déplacer cet équilibre vers la forme active (Robinson *et al.*, 2000; West et Stock, 2001). Les surfaces exposées lors de cette activation permettraient au régulateur de réponse d'acquérir la capacité de fixation à l'ADN ou permettraient l'apparition d'une activité enzymatique. Dans plusieurs cas, le domaine receveur inhibe l'activité du domaine effecteur. Le changement de structure suite à la phosphorylation permettrait de replacer les extrémités N et C terminales du RR, ce qui pourrait libérer les sites de reconnaissance de l'ADN du domaine effecteur.

Les RR ont en général une activité autophosphatase qui limite la durée de vie du RR à l'état phosphorylé (West et Stock, 2001). Cette activité permet un retour rapide à l'état non activé, ce qui permet une réponse rapide aux variations des conditions environnementales.

III.3) Exemples de mécanismes de transfert et de régulations.

III.3.1) Système EnvZ et OmpR.

Le système EnvZ/OmpR est un des systèmes à deux composants les plus étudiés chez *E. coli*. Il permet de détecter des variations de l'osmolarité du milieu et de modifier en conséquence l'expression des gènes codant les porines OmpF et OmpC (cf. paragraphe I.1) (pour revue : Pratt *et al.*, 1996; Egger *et al.*, 1997). En effet, le système d'osmorégulation d'*E. coli* lui permet de survivre dans des milieux à faible concentration saline comme l'eau et dans des milieux à forte concentration saline comme l'intestin humain (pour une revue sur les effets de l'osmolarité chez les bactéries, voir Wood, 1999). Lorsque l'osmolarité du milieu est faible, la porine OmpF est présente à une plus forte concentration dans la membrane externe que la porine OmpC. En effet, comme la taille des pores est plus élevée pour OmpF que pour OmpC, la surproduction de OmpF en conditions de faible osmolarité est un avantage pour permettre l'entrée des nutriments dans la cellule.

III.3.1.1) Mécanisme permettant la régulation de la concentration de OmpR-P.

Le domaine périplasmique de l'histidine kinase EnvZ permet de détecter les variations de l'osmolarité du milieu. Le signal qui lie l'activité de EnvZ à l'osmolarité du milieu n'est pas connu. Lorsque l'osmolarité du milieu augmente, le domaine senseur active plus fortement le domaine kinase et ainsi entraîne une autophosphorylation de l'histidine essentielle (His 243) au niveau de ce domaine (Roberts *et al.*, 1994). Le phosphate est ensuite transféré sur l'acide aspartique essentiel (Asp 55) du régulateur de réponse OmpR (Forst *et al.*, 1989; Aiba *et al.*, 1989; Aiba *et al.*, 1989). La concentration de OmpR-P dans la cellule augmente donc en fonction de l'activité de EnvZ. OmpR-P se fixe sur l'ADN en amont des promoteurs des gènes *ompF* et *ompC* (Huang et Igo, 1996), ce qui permet l'expression différentielle des porines OmpF et OmpC en fonction de l'osmolarité. Le retour à un taux basal de OmpR-P est permis grâce à l'activité phosphatase de EnvZ (pour revue : Egger *et al.*, 1997).

Récemment, ce modèle a été précisé (Fig. 38 et 39) (Yoshida *et al.*, 2002). Il existerait un équilibre entre plusieurs formes de la protéine EnvZ, une forme favorable à l'activité kinase et une forme favorable à l'activité phosphatase. La forme phosphatase serait majoritaire à faible osmolarité et la forme kinase serait majoritaire à forte osmolarité. Comme OmpR et OmpR-P

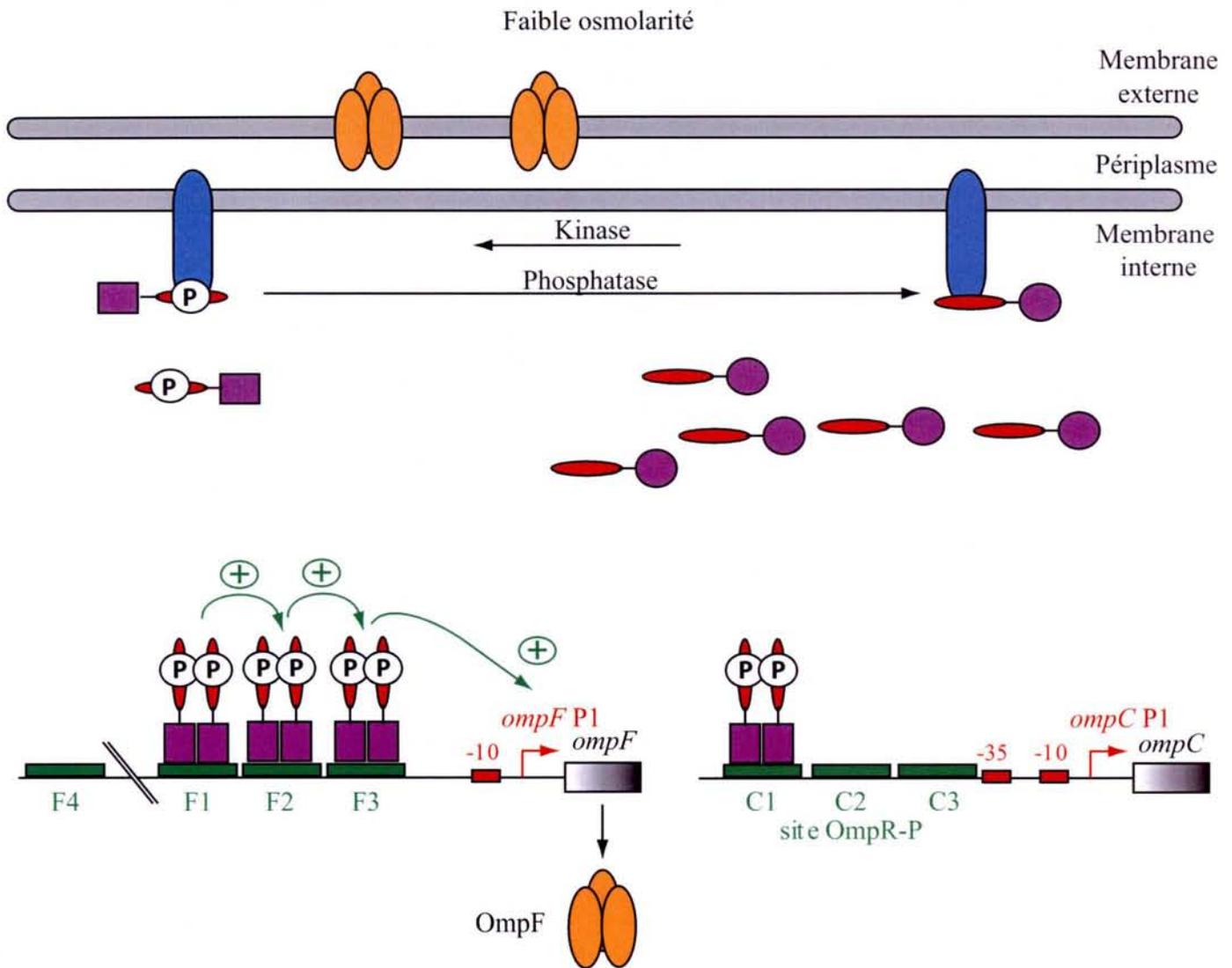


Figure 38 : Modèle de régulation de l'expression des porines OmpF et OmpC chez *E. coli* (d'après Egger *et al.*, 1997 ; Yoshida *et al.*, 2002).

A faible osmolarité, la porine majoritairement présente dans la membrane externe est OmpF (en orange). L'histidine kinase EnvZ est représentée en bleu. Elle peut exister sous deux formes différentes : une forme phosphatase et une forme kinase, qui sont en équilibre. A faible osmolarité, l'équilibre est déplacé vers la forme phosphatase. La quantité de OmpR-P est donc faible dans la cellule. Le domaine receveur de OmpR est représenté en rouge. L'état phosphorylé du domaine receveur de OmpR est indiqué par un P cerclé. Le domaine effecteur est représenté par des ronds violets lorsque le RR est inactif (domaine receveur non phosphorylé) et par des carrés violets lorsqu'il est actif (domaine receveur phosphorylé). Sous cette dernière forme, il peut se fixer à l'ADN sous forme de dimère sur des opérateurs situés en amont des promoteurs des gènes *ompF* et *ompC* (représentés en vert). Les sites d'initiation de la transcription sont indiqués, ainsi que les boîtes -10 et -35 (rouge).

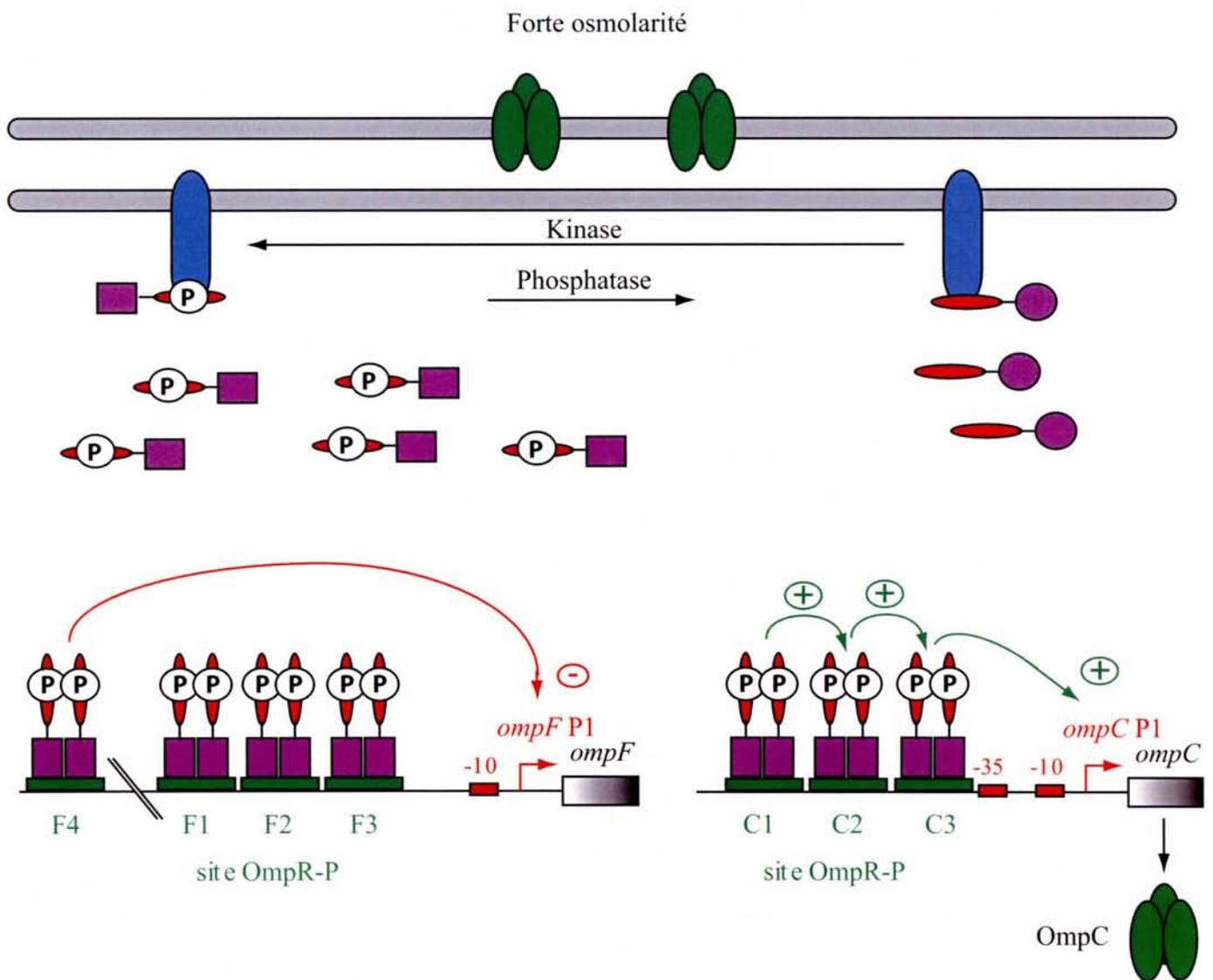


Figure 39 : Modèle de régulation de l'expression des porines OmpF et OmpC chez *E. coli* (d'après Egger *et al.*, 1997 ; Yoshida *et al.*, 2002).

A forte osmolarité, la porine majoritairement présente dans la membrane externe est OmpC (en vert) et l'équilibre des formes de la protéine EnvZ est déplacé vers la forme kinase. La quantité de OmpR-P est donc élevée dans la cellule. Un dimère de OmpR-P peut se fixer à l'ADN sur des opérateurs de plus faible affinité, situés en amont des promoteurs des gènes *ompF* et *ompC* (représentés en vert). Les sites d'initiation de la transcription sont indiqués, ainsi que les boîtes -10 et -35 (rouge).

pourraient se fixer sur EnvZ avec quasiment la même affinité, à faible osmolarité, le taux de OmpR-P serait faible. A forte osmolarité, l'activité kinase majoritaire conduirait à l'augmentation du taux de OmpR-P (Yoshida *et al.*, 2002).

III.3.1.2) Sites de fixation de OmpR-P sur l'ADN et régulation de la synthèse des porines.

La protéine OmpR-P sous forme dimérique peut se fixer sur plusieurs sites présents en amont des promoteurs des gènes *ompF* et *ompC* (Fig. 38 et 39). Il existe quatre sites de fixation de OmpR-P en amont du promoteur *ompF* P1 (F1, F2, F3 et F4) et trois en amont du promoteur *ompC* P1 (C1, C2 et C3) (Huang et Igo, 1996; Head *et al.*, 1998). Les sites F1 à F3 sont positionnés de -107 à -39 par rapport au site d'initiation de la transcription *ompF* P1, alors que le site F4 est positionné très en amont (positions -351 à -384). Les sites C1 à C3 sont respectivement positionnés aux positions -107 à -40 par rapport au site d'initiation *ompC* P1. Les sites F1 et C1 sont ceux ayant la plus forte affinité pour OmpR-P (Head *et al.*, 1998). La fixation de OmpR-P au niveau du site F1 permettrait la fixation coopérative de OmpR-P au niveau des sites F2 et F3 (Bergstrom *et al.*, 1998). La fixation au site F3 permettrait l'activation de l'initiation de la transcription au niveau du promoteur *ompF* P1 par interaction avec le domaine α -CTD de l'ARN pol et ainsi la production de la porine OmpF (Bowrin *et al.*, 1994). Le site F4 a beaucoup moins d'affinité pour OmpR-P et n'est occupé que lorsque la concentration de OmpR-P est très élevée. De la même manière, la fixation coopérative de OmpR-P aux sites C2 et C3 nécessite une concentration plus élevée en OmpR-P que pour la fixation coopérative aux sites F2 et F3 (Bergstrom *et al.*, 1998).

La présence de plusieurs sites de fixation avec des affinités différentes pour OmpR-P en amont des promoteurs des gènes *ompC* et *ompF* est responsable de l'expression différentielle des deux porines (pour revue : Egger *et al.*, 1997). En effet, à faible osmolarité, le taux de OmpR-P est faible et seuls les sites F1, F2, F3 et C1 sont occupés (Fig. 38). Dans ce cas, l'initiation de la transcription à partir du promoteur *ompF* P1 est activée et la protéine majoritaire présente dans la membrane est OmpF. Lorsque l'osmolarité augmente, le taux de OmpR-P augmente et les sites C2 et C3 sont occupés (Fig. 39). La transcription à partir du promoteur *ompC* P1 est activée, ce qui permet la production de la porine OmpC. A l'opposé, OmpR-P peut se fixer sur le site F4 et inhiber l'initiation de la transcription à partir du promoteur *ompF* P1 par formation d'une boucle d'ADN. Dans ce cas, la porine majoritaire dans la membrane externe est OmpC.

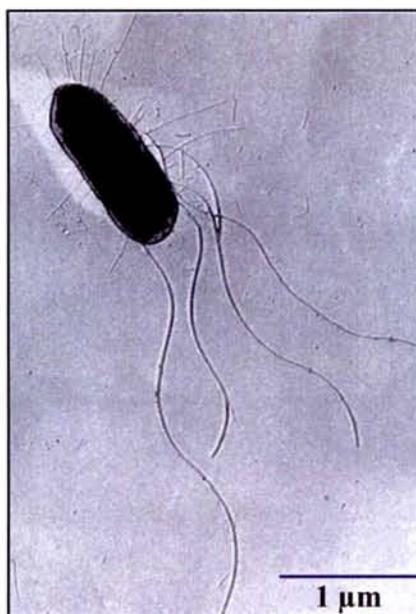
D'autres facteurs protéiques peuvent influencer directement ou indirectement l'expression des gènes *ompF* et *ompC*, comme par exemple, H-NS, CRP-AMPc, IHF et Lrp (Liu et Ferenci, 2001). Un ARN antisens *micF*, transcrit dans la direction opposée au gène *ompC*, permet d'inhiber la traduction de l'ARN *ompF* (Aiba *et al.*, 1987; Goosen et van de Putte, 1995; Suzuki *et al.*, 1996; Delihias et Forst, 2001). L'expression de cet ARN régulateur serait liée à l'occupation des sites de fixation de OmpR-P C1, C2 et C3 présents en amont du promoteur *ompC* P1. L'occupation de ces sites est en effet essentielle à la transcription de *micF* en fonction de l'osmolarité du milieu (Delihias et Forst, 2001).

L'effet de CRP-AMPc sur l'expression des porines OmpC et OmpF serait lié à la régulation de la transcription de l'opéron *ompB* codant EnvZ et OmpR. En effet, CRP-AMPc influence l'initiation de la transcription à partir des multiples promoteurs de *ompB* mais le lien entre la transcription de *ompB* et taux de phosphorylation de OmpR est incertain (Huang *et al.*, 1992). Ces régulations permettent l'expression contrôlée des porines OmpF et OmpC pour permettre un taux de diffusion optimale des nutriments à travers la membrane externe en fonction des conditions environnementales.

III.3.2) Le chimiotactisme chez *E. coli*.

Le chimiotactisme chez les bactéries est lié au déplacement en trois dimensions de la cellule, en général dans le but de se diriger vers les gradients croissants de sources de nutriments et de s'écarter des gradients croissants de sources de produits toxiques (se diriger vers des gradients décroissants de produits toxiques). Les éléments cellulaires qui permettent la locomotion de la bactérie sont les flagelles. Chez *E. coli*, il existe six à huit flagelles répartis de façon aléatoire à la surface de la cellule (Fig. 40). La rotation de ces flagelles permet à *E. coli* deux types de mouvements. Des périodes de mouvement rectiligne sont entrecoupées de retournements qui permettent la réorientation de la cellule. Ces deux types de mouvements permettent un déplacement net vers les sources de nutriments ou au contraire une réorientation de la cellule en présence d'une source de produits toxiques. Ces déplacements sont gérés par le système à deux composants CheA/CheY. Après une description des flagelles, je décrirai le fonctionnement de ce système à deux composants.

A



B

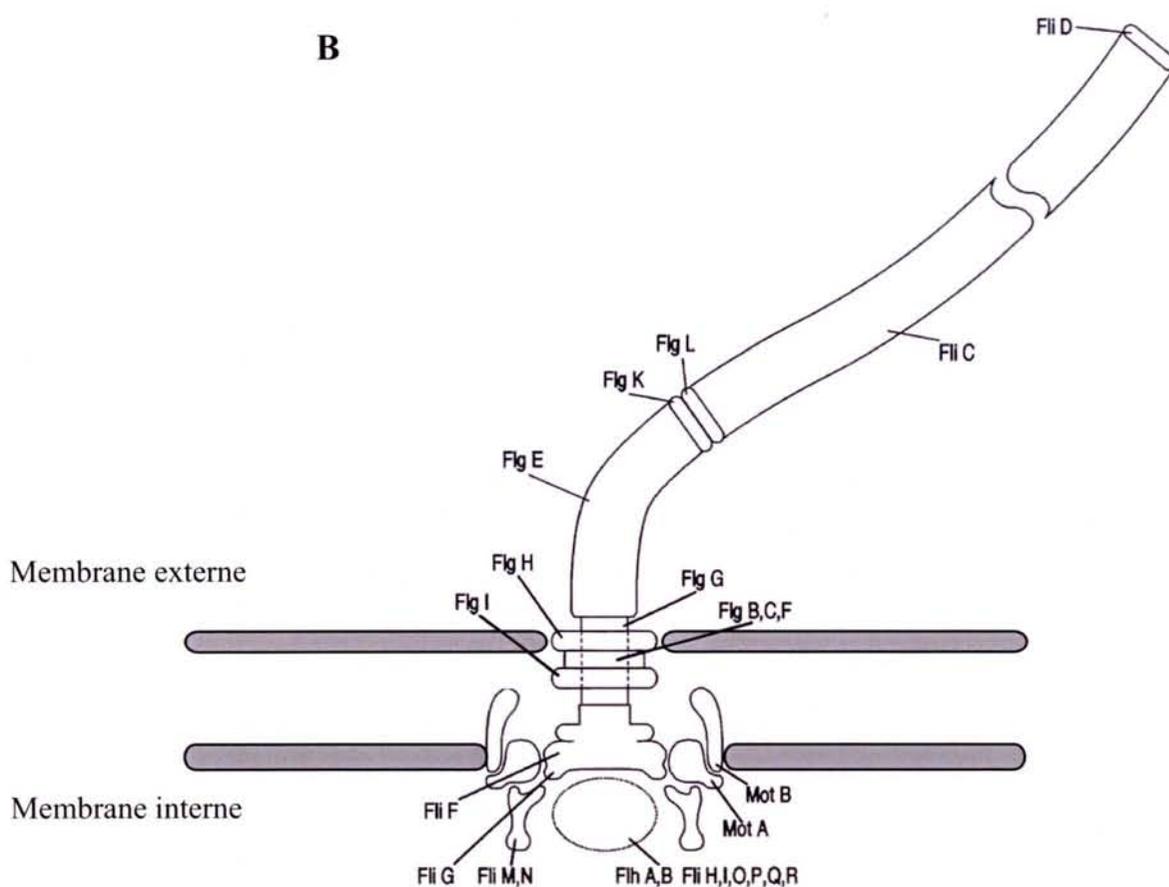


Figure 40 : Les flagelles d'*E. coli* (d'après McNab, 1996 et Berg, 2003).

A) Etude par microscopie électronique d'une cellule d'*E. coli* montrant les flagelles. La barre représente 1 μm.

B) Représentation schématique du flagelle. Les différents constituants du flagelle sont indiqués. Les membranes externes et internes sont représentées en gris.

III.3.2.1) Les flagelles.

Le flagelle est un tube fin dont la taille peut varier (jusqu'à dix fois le diamètre de la cellule d'*E. coli*). Il est constitué d'un grand nombre de protéines (Fig. 40 A) (pour revue : Macnab, 1996; Berg, 2003). Le flagelle peut tourner dans le sens des aiguilles d'une montre (CW : clockwise rotation) et dans le sens inverse (CCW : counterclockwise rotation). Lors du déplacement de la cellule, la nage rectiligne est due à une rotation CCW des flagelles, alors que la phase de retournement (réorientation) est due à une rotation CW de un ou plusieurs flagelles. L'énergie de la rotation du flagelle est fournie par la force proton-motrice (Kojima et Blair, 2004). Une partie du flagelle est insérée dans les deux membranes (interne et externe) de la bactérie (Fig. 40 B) (pour revue : Macnab, 1996; Berg, 2003).

Les gènes (dont le nom est en général identique au nom de la protéine) codant les protéines de constitution du flagelle sont organisés de façon hiérarchique en trois classes (Berg, 2003; Soutourina et Bertin, 2003). Dans la classe I, l'opéron *flhDC* code un facteur de transcription tétramérique FlhD₂C₂ qui régule l'expression des gènes de classe II (Liu et Matsumura, 1994; Claret et Hughes, 2002). Huit opérons qui constituent la classe II, codent les constituants de la base du flagelle mais aussi des protéines de régulation de l'expression des gènes de classe III comme le facteur σ^{28} (FliA). Les gènes de classe III codent aussi certains constituants du flagelle (filament externe) comme FliC et FliD et les constituants du moteur moléculaire (comme MotA et MotB), et aussi les protéines impliquées dans le chimiotactisme.

La synthèse des flagelles dépend uniquement de la concentration du facteur de transcription FlhD₂C₂. En conséquence, cette concentration est fortement régulée en fonction des conditions environnementales notamment au niveau de la transcription de l'opéron *flhDC* (pour revue : Soutourina et Bertin, 2003). Le complexe CRP-AMPc permet d'activer l'initiation de la transcription de cet opéron (Soutourina *et al.*, 1999). En effet, CRP-AMPc se fixe à la position -71,5 par rapport au site +1 d'initiation de la transcription et interagit avec l'ARN pol (Soutourina *et al.*, 1999). En présence de glucose dans le milieu de culture, le taux de CRP-AMPc diminue et l'initiation de la transcription à partir du promoteur de l'opéron *flhDC* est diminuée. La synthèse des flagelles est donc supprimée en présence de glucose dans le milieu de culture (Berg, 2003; Adler et Templeton, 1967). D'autres régulateurs comme H-NS (Soutourina *et al.*, 1999) et OmpR-P (Shin et Park, 1995) affectent l'initiation de la transcription de façon négative à partir du promoteur de l'opéron *flhDC*.

Ces régulations permettent une expression optimale du régulateur FlhD₂C₂ en fonction des conditions environnementales. Ce facteur de transcription permet d'activer l'initiation de la transcription des gènes de classe II et notamment le gène *fliA* qui code le facteur $\sigma^{28(\text{FliA})}$ (Claret et Hughes, 2002). Le site de fixation de FlhD₂C₂ sur l'ADN chevauche la séquence de la boîte -35 du promoteur du gène *fliA* et pourrait interagir avec la sous-unité α de l'ARN pol (Claret et Hughes, 2002). Ce mode de fixation serait identique pour les autres gènes de la classe II. L'affinité des différents opérateurs pour FlhD₂C₂ semble expliquer l'expression temporelle des gènes de classe II (Claret et Hughes, 2002; Soutourina et Bertin, 2003). Le facteur FlhD₂C₂ pourrait être un facteur de transcription global. En effet, une approche par puce à ADN a montré que ce complexe régule l'expression d'autres gènes qui ne sont pas impliqués dans la synthèse des flagelles comme par exemple le transporteur du galactose (*mgIBAC*) ou la malate déshydrogénase du cycle de Krebs (*mdh*) (Pruss *et al.*, 2001).

En association avec l'apoenzyme, le facteur σ^{28} permet d'initier la transcription des gènes de classe III. L'activité de ce facteur σ est régulée par le facteur antisigma FlgM (cf. partie régulation de la transcription).

L'assemblage du flagelle dans les membranes de la bactérie est un phénomène complexe qui ne sera pas décrit ici (pour revue : Aldridge et Hughes, 2002; Kojima et Blair, 2004).

En conclusion, la synthèse des flagelles chez *E. coli*, qui dépend uniquement du facteur de transcription FlhD₂C₂, est régulée en fonction de l'environnement dans lequel évolue la bactérie. Un autre niveau de régulation est l'activité du moteur qui est régulée par le système à deux composants CheA/CheY.

III.3.2.2) Le système CheA/CheY.

Les protéines du chimiotactisme peuvent être réparties en quatre groupes. Le premier groupe représente les protéines qui permettent la reconnaissance du signal. Le deuxième groupe est représenté par les protéines qui transduisent le signal (excitation), alors que les protéines des troisième et quatrième groupes permettent un retour à un état non excité (adaptation et arrêt du signal) (pour revue : Stock et Surette, 1996; Manson *et al.*, 1998; Bourret et Stock, 2002; Szurmant et Ordal, 2004).

Les protéines du premier groupe sont les récepteurs membranaires qui permettent la reconnaissance du signal (pour revue : Stock et Surette, 1996; Manson *et al.*, 1998; Falke et Hazelbauer, 2001). Ces récepteurs, qui peuvent être de cinq types différents chez *E. coli*, sont

des homodimères formés même en absence de ligand (Manson *et al.*, 1998). Ils sont constitués d'une région périplasmique où peuvent se fixer les ligands, et d'une région cytoplasmique où peuvent se fixer les protéines de transmission du signal comme CheA (Fig. 41) (Stock et Surette, 1996; Bourret et Stock, 2002). En absence de ligand, la formation d'un complexe entre le récepteur et CheA, par l'intermédiaire de CheW, active l'autophosphorylation de cette dernière sur l'histidine conservée. CheW permet d'adapter l'HK CheA aux différents récepteurs pour conduire à une intégration des signaux délivrés par ces récepteurs (Bourret et Stock, 2002). Le phosphate est ensuite transmis sur un acide aspartique de CheY dont la structure est alors modifiée (Robinson *et al.*, 2000). CheY-P va alors interagir avec la base du moteur du flagelle, notamment avec la protéine FliM (Welch *et al.*, 1993). Cette interaction va changer le sens de rotation du flagelle, qui passe d'une rotation CCW à une rotation CW (Fig. 41) (Stock et Surette, 1996; Manson *et al.*, 1998). Dans ce cas, la bactérie va passer d'une période de mouvement rectiligne à une période de retournement qui va permettre une réorientation du sens de déplacement de la cellule.

Dans le cas de la fixation d'un ligand au niveau du domaine périplasmique du récepteur, une longue hélice qui relie le domaine périplasmique au domaine cytoplasmique permet la transmission du signal à l'intérieur de la bactérie (Falke et Hazelbauer, 2001). Chez *E. coli*, ce signal de fixation modifie l'autophosphorylation de CheA (Fig. 41) (pour revue : Falke et Hazelbauer, 2001). Dans le cas de la fixation d'un ligand « attractant » comme l'aspartate, le galactose ou la sérine au niveau de récepteurs spécifiques, respectivement, Tar, Trg ou Tsr, l'autophosphorylation de CheA est inhibée. De cette manière, le phosphate n'est pas transmis à CheY et la période de mouvement rectiligne est allongée par rapport à la période de retournement. Le déplacement net de la bactérie lui permet alors de se diriger vers cette source de nutriments. A l'inverse, en absence de ligands « attractants » ou en présence de produits toxiques, l'autophosphorylation de CheA augmente. Les périodes de retournement sont plus longues et la bactérie se réoriente plus souvent (pour revue : Stock et Surette, 1996; Silversmith et Bourret, 1999).

Pour revenir à l'état non excité, CheY possède une activité d'auto-déphosphorylation. La durée de vie de CheY-P est très faible (Stock et Surette, 1996; Bourret et Stock, 2002). De ce fait, pour que le flagelle change de sens de rotation, il faut une forte concentration en CheY-P et donc une très forte activation de CheA. De plus, la phosphatase CheZ peut faciliter la déphosphorylation de CheY (Zhao *et al.*, 2002). Son activité permet de diminuer rapidement la quantité de CheY-P (pour revue : Manson *et al.*, 1998; Falke et Hazelbauer, 2001).

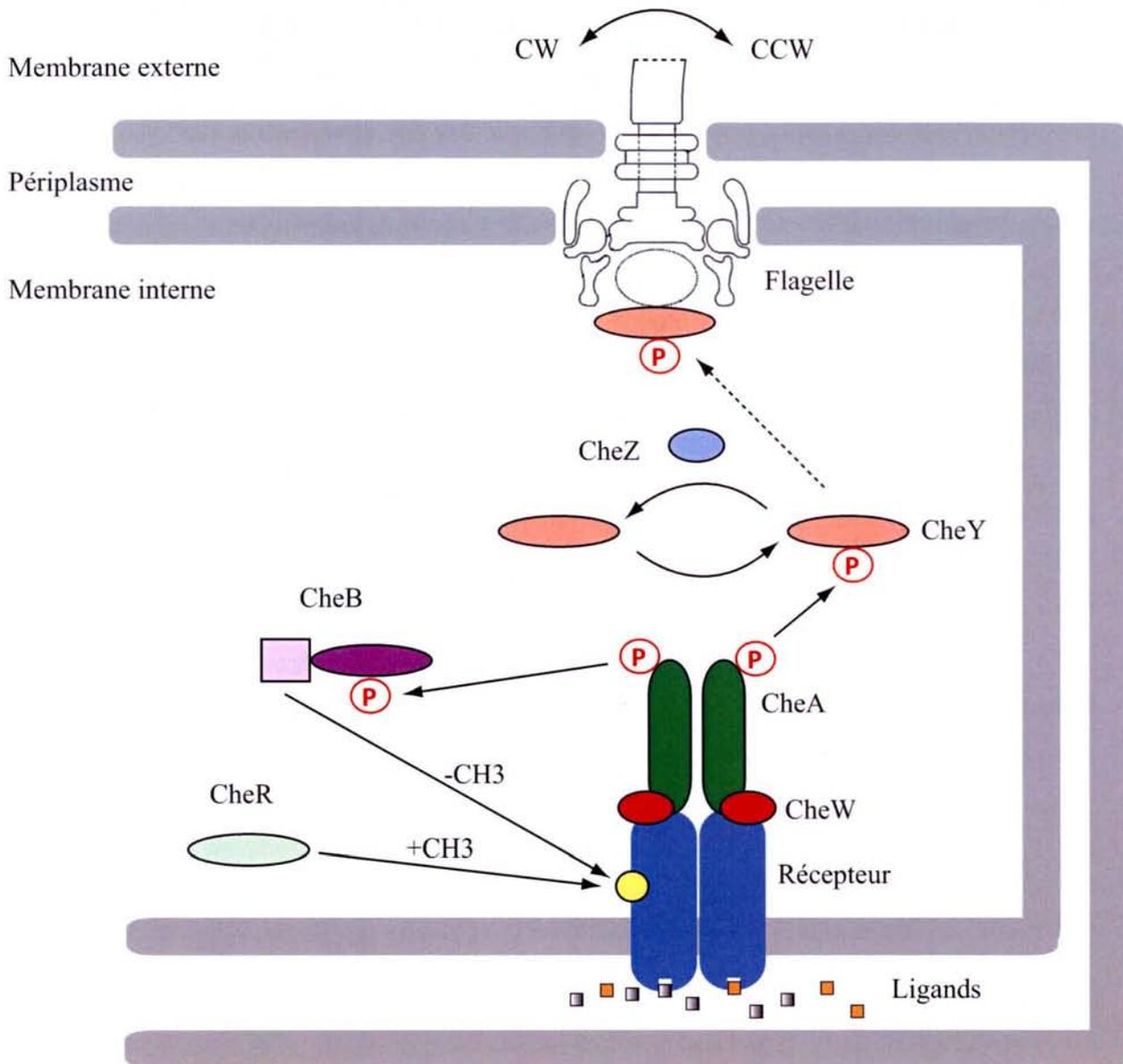


Figure 41 : Le chimiotactisme chez *E. coli* (d'après Stock et Surette, 1996 et Bourret et Stock, 2002).

Les protéines qui interviennent dans le chimiotactisme sont représentées par des couleurs différentes. Le récepteur membranaire est représenté en bleu et le site de méthylation par CheR (vert clair) est représenté en jaune. La protéine CheB (violet) qui est activée par la phosphorylation de CheA peut enlever le groupement méthyle fixé sur le récepteur (indiqué par -CH₃). L'adaptateur CheW (rouge) permet la fixation de CheA (vert foncé) sur le récepteur. La protéine CheY est représentée en orange. La structure de cette protéine est modifiée par la phosphorylation de CheA. CheY-P va ensuite interagir avec le moteur du flagelle. Le sens de rotation du flagelle passe de CCW (sens inverse des aiguilles d'une montre) à CW (sens des aiguilles d'une montre), ce qui change l'orientation du déplacement de la bactérie. CheY-P peut être déphosphorylée par la phosphatase CheZ (bleu). L'activité des récepteurs dépend de la présence de ligands dans l'espace périplasmique.

L'activité des récepteurs est régulée par deux protéines CheB et CheR (Fig. 41). CheR méthyle de façon constitutive le récepteur membranaire, alors que CheB qui, lors de la phosphorylation par CheA, devient active, peut enlever les groupements méthyles sur le récepteur (Lupas et Stock, 1989). En absence de molécules « attractantes », CheA s'autophosphoryle, ce qui active, entre autre, CheB. Le taux de méthylation du récepteur est faible, ce qui le rend très sensible à la présence d'agents « attractants » (Borkovich *et al.*, 1992). En présence de molécules « attractantes », l'autophosphorylation de CheA est inhibée par le récepteur, ce qui limite aussi l'activité de CheB. Le taux de méthylation du récepteur augmente, ce qui le rend moins sensible aux agents « attractants » (Borkovich *et al.*, 1992). Il faut plus de molécules « attractantes » pour entraîner l'inhibition de CheA. Ce système permet donc une adaptation à la concentration de molécules « attractantes », ce qui rend possible la détection de concentrations encore plus élevées de molécules « attractantes » (pour revue : Bourret et Stock, 2002).

CheY serait donc une protéine navette qui circule entre deux complexes membranaires : le complexe récepteur, CheW et CheA et le complexe du moteur du flagelle, pour réguler le déplacement de la bactérie en fonction des conditions environnementales.

Partie 3 : Présentation du sujet

Deux gènes codent des protéines de la famille des GAPDH chez *E. coli*.

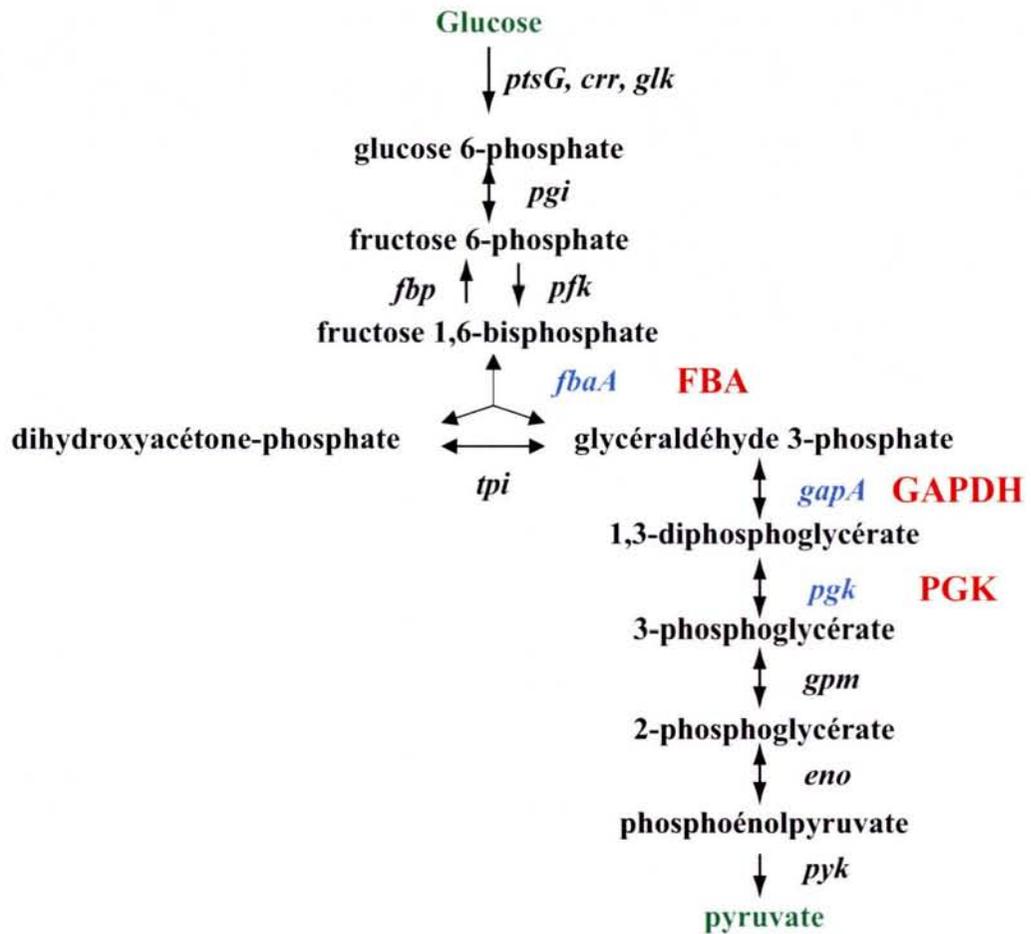
La glycolyse est la principale voie de dégradation des sources carbonées pour un grand nombre d'organismes allant des bactéries aux eucaryotes supérieurs. Cette suite de 9 réactions enzymatiques permet la formation de pyruvate à partir de glucose (Fig. 42). Les enzymes de la glycolyse notamment la glyceraldéhyde 3-phosphate déshydrogénase (GAPDH), la phosphoglycérate kinase (PGK) et la fructose 1,6-bisphosphate aldolase (FBA) sont fortement exprimées et l'organisation des gènes codant ces enzymes est différente suivant les organismes.

Chez la majorité des bactéries, le gène *gap* codant la GAPDH active dans la glycolyse est localisé dans un opéron avec d'autres gènes codant des enzymes impliquées dans la glycolyse. Il est le plus souvent localisé en amont du gène *pgk* codant la PGK, c'est par exemple le cas chez *Corynebacterium glutamicum* ou *Bacillus megaterium* (Eikmanns, 1992; Schlapfer et Zuber, 1992). Dans la majorité des bactéries à Gram⁺, l'opéron contenant les gènes *gap* et *pgk* contient aussi le gène *tpi* codant la triosephosphate isomérase (TIM) (Bardey, 2000).

En revanche, chez *E. coli*, deux gènes codent des protéines présentant de fortes similitudes avec les GAPDH. Ces deux gènes sont localisés à des loci différents sur le chromosome. Le gène *gapA* est localisé à 39,3 minutes (Thomson *et al.*, 1979; Branlant et Branlant, 1985) et le gène *gapB* (maintenant appelé *epd*) est localisé à 61,9 minutes (Alefounder *et al.*, 1989; Bardey, 2000). L'ORF *epd* qui code une protéine ayant une activité érythrose-4-phosphate déshydrogénase non phosphorylante (E4PDH) (Zhao *et al.*, 1995; Boschi-Muller *et al.*, 1997), convertissant l'érythrose 4-phosphate en 4-phosphoérythronate (Boschi-Muller *et al.*, 1997), est localisée en amont des ORF *pgk* et *fbxA* et cette organisation ne semble exister que chez les protéobactéries de la sous-division γ (Bardey, 2000).

Le gène *gapA* code la GAPDH active dans la glycolyse. Cette enzyme possède une plus forte similitude avec les GAPDH d'eucaryotes qu'avec beaucoup de GAPDH bactériennes (Branlant et Branlant, 1985). De ce fait, il a été proposé que le gène *gapA* ait été acquis par transfert horizontal de matériel génétique d'un ancêtre eucaryote vers un ancêtre des bactéries. Le gène de l'E4PDH aurait pu être généré par mutation d'un gène GAPDH ancestral (Branlant et Branlant, 1985; Martin *et al.*, 1993).

A



B

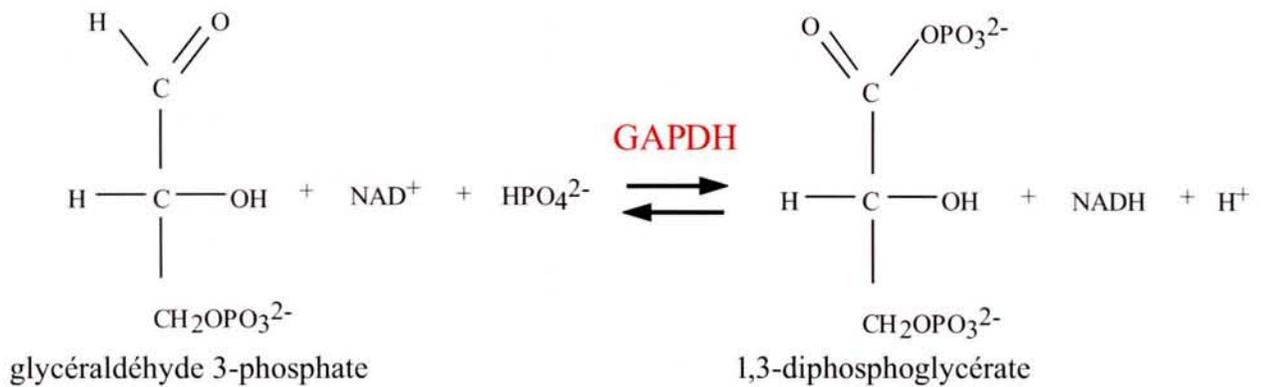


Figure 42 : Représentation de la glycolyse (A) et de la réaction catalysée par les GAPDH phosphorylantes (B).

Les gènes codant les enzymes de la glycolyse, ainsi que les enzymes FBA, GAPDH et PGK sont indiqués.

I.) Le gène *gapA* d'*E. coli*.

I.1) Présentation de l'enzyme : la GAPDH (EC 1.2.1.13).

Les glycéraldéhyde 3-phosphate déshydrogénases sont des enzymes ubiquitaires et sont réparties en trois grandes classes suivant leur mécanisme réactionnel et leur rôle physiologique. Les GAPDH glycolytiques catalysent la phosphorylation oxydative du D-glycéraldéhyde 3-phosphate (G3P) (produit de la réaction réalisée par la fructose-1,6-bisphosphate aldolase) en 1,3-diphosphoglycérate. Le cofacteur de la réaction est le NAD^+ (Fig. 42 B) (Harris et Waters, 1976). La deuxième famille de GAPDH catalyse préférentiellement la réaction inverse des GAPDH glycolytiques c'est-à-dire la déphosphorylation réductrice du 1,3-diphosphoglycérate en D-glycéraldéhyde 3-phosphate. De plus, ces enzymes utilisent plutôt le NADP^+ comme cofacteur. Ces enzymes seraient donc impliquées dans la réaction inverse de la glycolyse : la néoglucogenèse (Koksharova *et al.*, 1998). La troisième famille de GAPDH est distincte des deux autres. Ces enzymes sont non phosphorylantes et utilisent le NADP^+ comme cofacteur. Elles catalysent la conversion du D-glycéraldéhyde 3-phosphate en 3-phosphoglycérate (Iglesias et Losada, 1988).

La participation à la glycolyse est la fonction principale de la GAPDH phosphorylante mais d'autres rôles ont été postulés pour cette enzyme (pour revue : (Sirover, 1996)). Ainsi, chez certains streptocoques, des formes de GAPDH ancrées dans la membrane externe semblent exister et être impliquées dans la pathogénicité des cellules (Pancholi et Fischetti, 1992).

La GAPDH bactérienne est une enzyme homotétramérique, le poids moléculaire d'un monomère est de 36 kDa (330 acides aminés). Un monomère est caractérisé par la présence de deux domaines distincts : un domaine de fixation du NAD^+ , appelé domaine 'Rossmann fold' et un domaine portant les acides aminés catalytiques (cystéine 149 et histidine 176), ainsi que les sites de fixation du G3P et du phosphate inorganique. La structure tridimensionnelle de la GAPDH a été déterminée pour plusieurs organismes, en particulier, pour *E. coli* (Fig. 43) (Duee *et al.*, 1996).

Le mécanisme réactionnel de la GAPDH peut être décomposé en deux étapes (Fig. 44). La première étape est une réaction d'oxydo-réduction, qui conduit à la formation d'un intermédiaire thioacylenzyme. La cystéine essentielle (C149) réalise une attaque nucléophile sur le carbone de l'aldéhyde du G3P. La réactivité de cette cystéine est augmentée à pH physiologique par la formation d'une paire d'ions avec une histidine 176. La formation de

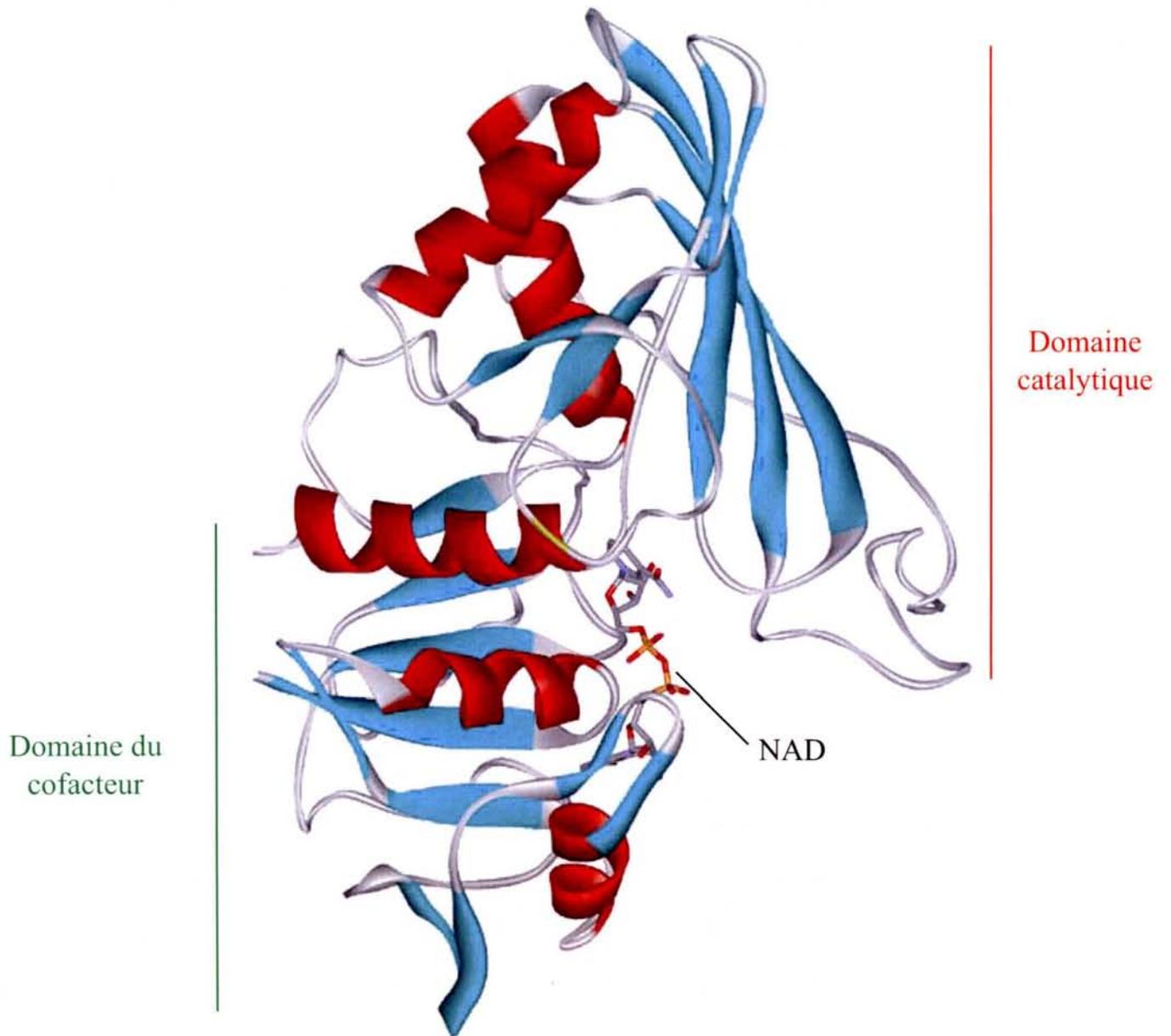
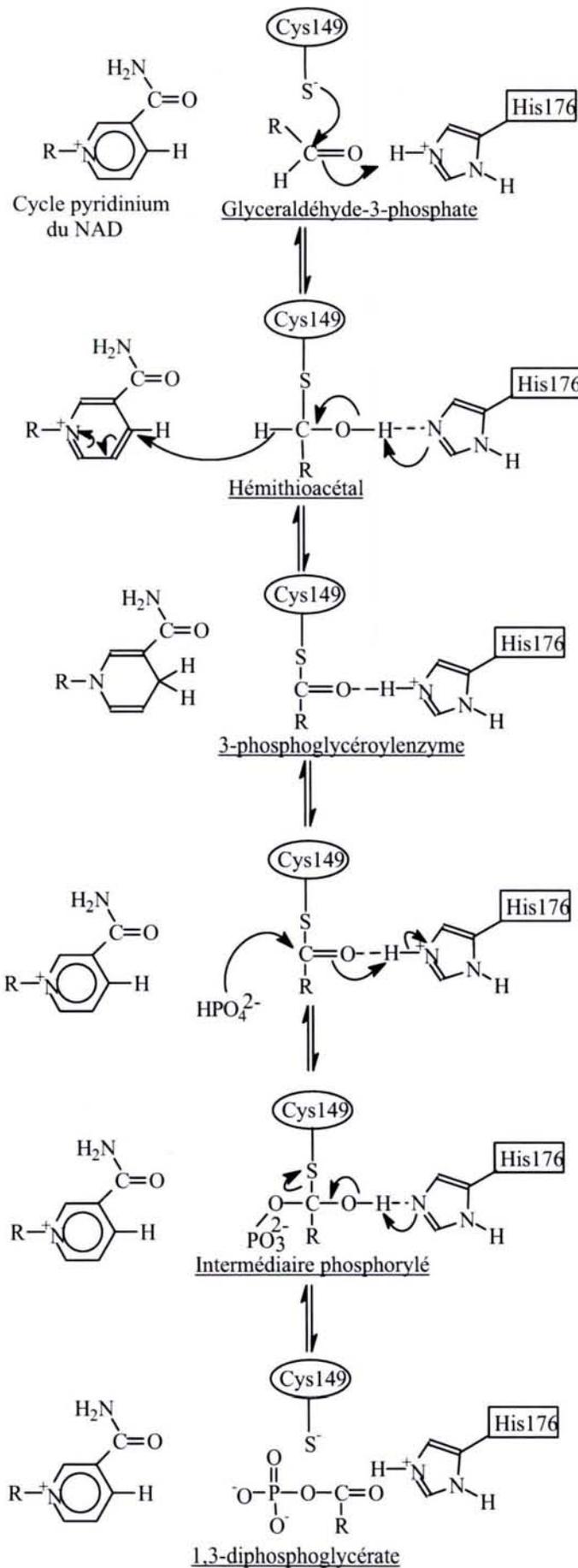


Figure 43 : Structure tertiaire du monomère de la GAPDH d'*E. coli* (d'après Yun *et al.*, 2000). Le monomère se structure en deux domaines différents : le domaine de fixation du cofacteur et le domaine catalytique. Les hélices α sont indiquées en rouge et les brins β en bleu. La molécule de NAD est indiquée.



La Cys 149, dont la réactivité est augmentée par le biais d'une paire d'ions avec l'His 176 protonnée, attaque le centre électrophile du G3P.

La formation de l'hémithioacétal est assistée par l'His 176 protonnée jouant le rôle d'acide.

(R = $\text{CHOH-CH}_2\text{OPO}_3^{2-}$)

L'hémithioacétal est stabilisé par liaison H avec l'His 176 protonnée.

L'His 176 jouant le rôle base assiste la réaction d'oxydoréduction, i.e. le transfert d'hydrure de l'hémithioacétal au pyridinium qui est ainsi réduit de NAD^+ à NADH.

L'acylenzyme est stabilisé par liaison H avec l'His 176 protonnée.

L'étape d'isomérisation de l'acylenzyme se fait après le remplacement du NADH par un NAD^+ .

L'attaque nucléophile d'un phosphate (phosphorolyse) est assistée par l'His 176 protonnée jouant le rôle d'acide.

Le relargage du 1,3-dPG est assisté par l'His 176 jouant le rôle de base.

L'enzyme se retrouve sous forme holoenzyme.

Figure 44 : Représentation du mécanisme réactionnel proposé pour les GAPDH phosphorylantes (d'après Segal et Boyer, 1953).

l'intermédiaire thioacylenzyme est permise grâce au transfert d'hydrure de l'intermédiaire hémithioacétal vers le groupement nicotinamide du cofacteur NAD^+ , qui est alors réduit. Ce transfert est facilité par le rôle de base joué par l'histidine 176. La deuxième étape est une phosphorylation qui permet de relarguer le produit formé : le 1,3-diphosphoglycérate (Fig. 44).

I.2) Transcription du gène *gapA* codant la GAPDH d'*E. coli*.

Ce gène est très fortement transcrit et l'étude approfondie de la région promotrice par B. Charpentier au laboratoire a permis de montrer que la complexité de cette région était responsable de la production de GAPDH à un taux quasiment constant dans différentes conditions de croissance d'*E. coli* (Charpentier, 1994; Charpentier et Branlant, 1994). En effet, le gène *gapA* est sous le contrôle de quatre promoteurs régulés de façon différente (Fig. 45) (Charpentier, 1994; Charpentier et Branlant, 1994). Il s'agit donc d'un système multipromoteurs d'une complexité peu courante chez les bactéries et donc intéressant à étudier. En dépit de cet intérêt, les informations disponibles sur cette région étaient encore très parcellaires lorsque j'ai débuté ma thèse.

Le promoteur P1, qui correspond au site majoritaire d'initiation de la transcription est extrêmement actif en phase exponentielle de croissance (Charpentier et Branlant, 1994). Il possède des boîtes -10 et -35 reconnues par l'holoenzyme $\text{E}\sigma^{70}$. Un couple de boîtes -10 et -35 avait été suggéré pour ce promoteur par analyse de séquence et identification du site de démarrage de la transcription : la séquence de la boîte -10 potentielle proposée était TAATTT et TTGACG celle de la boîte -35 (Fig. 45). La distance entre ces deux boîtes était sous-optimale (16 pb). L'utilisation de ce couple de boîte n'avait pas été démontrée clairement par des expériences de mutagenèse dirigée et un autre couple chevauchant correspondant à un promoteur étendu pouvait être le promoteur actif. Un couple de boîtes -10 P'1 et -35 P'1 avait aussi été observé sur la base de la séquence, mais ce couple s'était avéré inefficace vraisemblablement du fait de la distance interboîte trop importante (19 pb). Il restait donc à discriminer entre les deux couples possibles cités ci-dessus. Le taux de transcrits initiés au promoteur *gapA* P1 est augmenté lorsque du glucose est présent dans le milieu de culture et cette augmentation est dépendante de l'expression de la protéine EIIBC^{Glc} (Charpentier *et al.*, 1998). Des sites potentiels de fixation pour la protéine Fis avaient été détectés en amont du promoteur P1. Ces sites potentiels sont centrés aux positions -62 ou -68 par rapport au site d'initiation P1 (Charpentier, 1994). Leur rôle n'avait pas été testé.

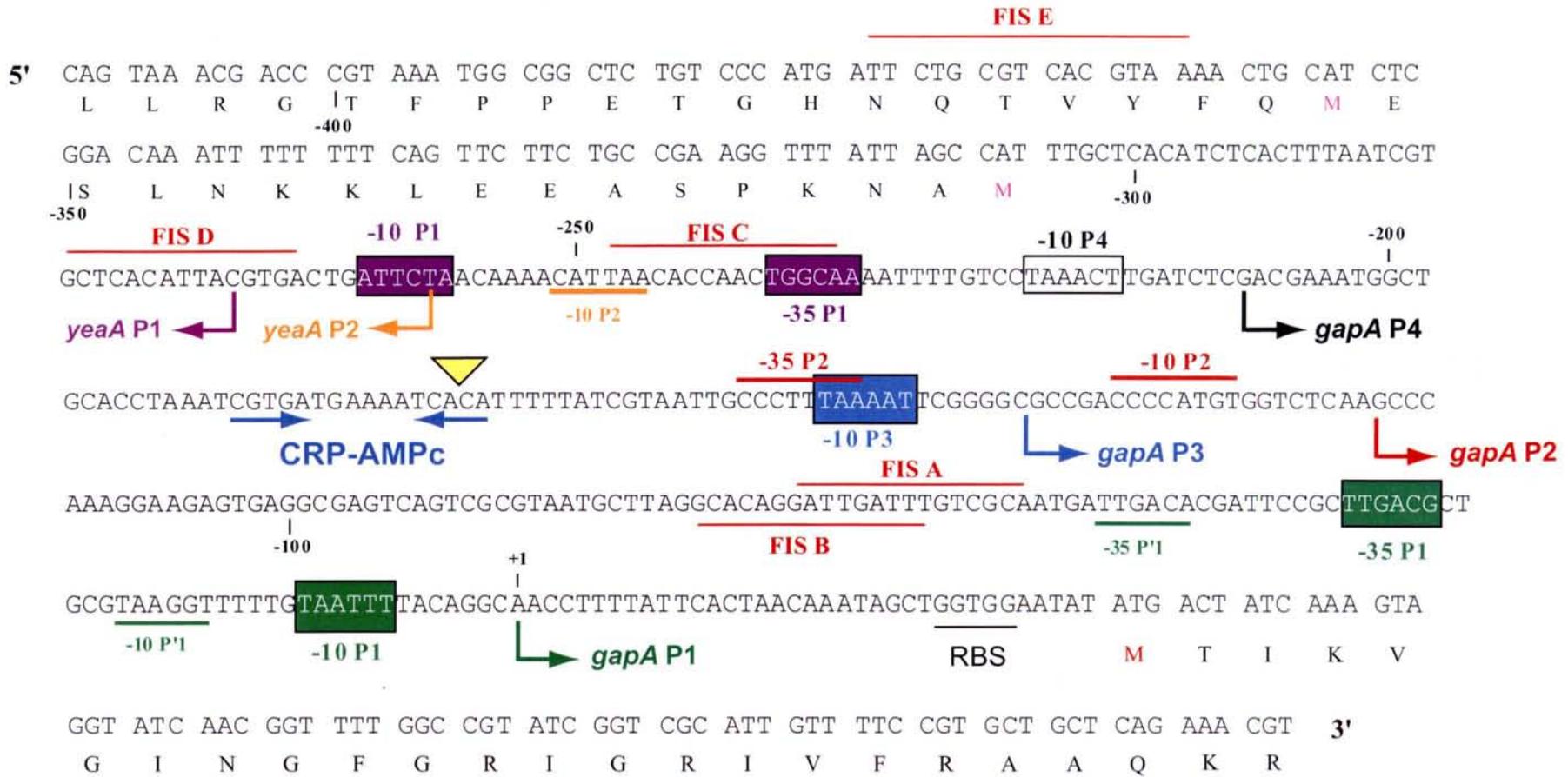


Figure 45 : Région régulant la transcription des gènes *gapA* et *yeaA* chez *E. coli* (d'après Charpentier et Branlant, 1994 ; Charpentier, 1994).

Les 4 sites d'initiation de la transcription du gène *gapA* sont indiqués par des flèches de couleurs différentes : P1 en vert, P2 en rouge, P3 en bleu et P4 en noir. Les boîtes -10 et -35 des promoteurs permettant l'initiation aux sites P1, P3 et P4 sont encadrées et les boîtes -10 et -35 dirigeant la transcription au site P2 sont soulignées. Les boîtes -10 et -35 du promoteur P'1 sont soulignées en vert. Les sites potentiels de fixation de la protéine FIS (A, B, C, D et E) sont indiqués par des traits rouges. Le centre d'une forte courbure de l'ADN est représenté par un triangle jaune, il est localisé au niveau du site de fixation du complexe CRP-AMPc activant l'initiation de la transcription au site P3. La séquence de l'ORF *yeaA* est indiquée, ainsi que les promoteurs *yeaA* P1 et P2 permettant la transcription de cette ORF. Les boîtes -10 P1, -35 P1, -10 P2 et les sites d'initiation de la transcription (représentés par des flèches) sont indiqués en violet et orange. RBS : site de fixation du ribosome.

Le promoteur P2 possède des boîtes -10 et -35 reconnues par l'holoenzyme $E\sigma^{32}$, ce qui permet une expression du gène *gapA* en réponse à un choc thermique (Fig. 45) (Charpentier, 1994). L'expression de la GAPDH est donc maintenue en conditions de choc thermique.

L'initiation de la transcription aux promoteurs P3 et P4 est faible en phase exponentielle de croissance (P4 étant le plus faible). L'utilisation du promoteur P3 est activée par le complexe CRP-AMPC, dont le site de fixation est centré à la position -40,5 (cette distance est compatible avec une activation par un facteur de classe II). L'utilisation du promoteur *gapA* P3 est donc soumise à la répression catabolique (Charpentier et Branlant, 1994). De plus, une forte courbure intrinsèque de l'ADN, qui est présente en amont de ce promoteur *gapA* P3, pourrait influencer son efficacité mais ce rôle n'avait pas clairement été mis en évidence. Enfin, le promoteur P3 ne possède pas de séquence ressemblant à une séquence de boîte -35. Le promoteur P4 est très faible *in vivo* et a été peu étudié.

D'autres sites potentiels de fixation pour la protéine Fis avaient été détectés en amont du promoteur P4 (Fig. 45). Ces sites pourraient chevaucher les séquences des boîtes -10 et -35 ainsi que les sites d'initiation de la transcription de deux promoteurs potentiels P1 et P2 qui permettraient la transcription de l'ORF *yeaA* sur le brin opposé à celui de *gapA*. B. Charpentier avait en effet montré que des transcrits étaient initiés à partir de ces promoteurs et que des complexes RPo étaient formés au niveau de leurs boîtes -10 (Charpentier, 1994).

L'ORF *yeaA* code une protéine qui possède une activité MsrB permettant la réduction des méthionines oxydées (Grimaud *et al.*, 2001). Des entités oxygénées réactives sont produites lors du métabolisme cellulaire ou lors de stress oxydants. Elles peuvent modifier de nombreuses cibles dont les protéines, notamment l'acide aminé méthionine. Les protéines Msr (méthionine sulfoxyde réductase) catalysent la réduction des méthionines sulfoxydes (MetSO) incluses au sein des protéines en méthionine (pour revue : (Weissbach *et al.*, 2002)). Deux isomères de la fonction sulfoxyde de la MetSO peuvent exister : l'isomère R et l'isomère S. La MsrA catalyse la réduction de l'isomère S, alors que la MsrB catalyse la réduction de l'isomère R (Sharov *et al.*, 1999; Moskovitz *et al.*, 2000; Grimaud *et al.*, 2001). Les mécanismes réactionnels ainsi que les structures de ces deux familles d'enzymes ont été déterminés, notamment pour les enzymes de *Neisseria meningitidis* et *E. coli* au laboratoire par l'équipe Enzymologie Moléculaire (Neiers *et al.*, 2004; Antoine *et al.*, 2003; Kauffmann *et al.*, 2002; Olry *et al.*, 2002; Boschi-Muller *et al.*, 2001; Tete-Favier *et al.*, 2000; Boschi-Muller *et al.*, 2000; Tete-Favier *et al.*, 2000).

La protéine MsrA d'*E. coli* est codée par le gène *msrA*, localisé à une position différente de celle du gène *yeaA* sur le chromosome d'*E. coli*. Ces protéines sont sans doute impliquées dans la résistance au stress oxydant lié par exemple à la production d'entités oxygénées réactives par les cellules immunitaires lors de l'infection bactérienne d'un organisme vertébré. La région inter-ORF *yeaA-gapA* semblait donc contenir 6 séquences promotrices et beaucoup d'éléments de régulation susceptibles de réguler l'un, l'autre ou les deux gènes qui la borde. Seules des études approfondies mais non complètes du promoteur *gapA* P1 avaient été réalisées dans cette région inter-ORF.

II.) L'opéron *epd-pgk-fbaA* d'*E. coli*.

Comme nous l'avons déjà mentionné, l'ORF *epd* code une protéine ayant une activité érythrose 4-phosphate déshydrogénase non phosphorylante (E4PDH) dont la structure 3D est très similaire à celle des GAPDH (Kauffmann *et al.*, résultats non publiés) (Zhao *et al.*, 1995; Boschi-Muller *et al.*, 1997; Fillingner *et al.*, 2000). Cette enzyme réalise l'oxydation de l'érythrose 4-phosphate (substrat possédant un groupement CHOH et donc un carbone asymétrique de plus que le G3P) en 4-phosphoérythronate et possède aussi une faible activité GAPDH phosphorylante (Boschi-Muller *et al.*, 1997). Il a été proposé que cette enzyme soit impliquée dans la biosynthèse du pyridoxal 5'-phosphate (Yang *et al.*, 1998; Zhao *et al.*, 1995). Cependant, des souches où les gènes *gapA* et *epd* avaient été éliminés pouvaient se développer sur un milieu ne contenant pas de pyridoxal, alors que ce composé est essentiel à la croissance de la bactérie (Della-Seta *et al.*, 1997). L'E4PDH ne serait donc pas essentielle à la voie de synthèse du pyridoxal. Le mécanisme réactionnel des E4PDH est très similaire à celui des GAPDH mis à part le fait que c'est une molécule d'eau qui permet le relargage du produit et non un phosphate.

Au laboratoire, V. Bardey, N. Robas et B. Charpentier ont étudié l'expression du tandem d'ORF *epd-pgk* d'*E. coli*. Le gène *epd* est transcrit à partir d'un promoteur unique, *gapB* P0, présentant des propriétés particulières (Fig. 46) (Charpentier, 1994; Charpentier *et al.*, 1998). Un site de fixation pour le complexe CRP-AMPc est positionné en amont du promoteur *gapB* P0, ainsi qu'un site de fixation pour la protéine Cra (Saier et Ramseier, 1996). De plus, le taux de transcrits initiés à ce promoteur augmente en présence de glucose dans le milieu de culture et cette augmentation est dépendante de la présence du transporteur spécifique du glucose EIIBC^{Glc} (Charpentier *et al.*, 1998; Bardey, 2000). Le complexe CRP-AMPc permettrait

d'augmenter légèrement l'efficacité du promoteur *gapB* P0 de façon directe mais l'effet majoritaire serait indirect (Charpentier *et al.*, 1998; Bardey, 2000). En effet, l'initiation de la transcription du gène *ptsG* qui code EIIBC^{Glc} est strictement dépendante du complexe CRP-AMPc (Plumbridge, 1998b). Le gène *pgk* serait transcrit à partir de deux promoteurs : le promoteur *gapB* P0 situé en amont du gène *epd* et un promoteur situé en amont du gène *pgk* encore à définir (Nellemann *et al.*, 1989; Charpentier *et al.*, 1998). Malgré la transcription efficace de l'ORF *epd*, la protéine E4PDH est exprimée à un taux très faible dans la cellule du fait du clivage de son ARN messenger au sein du transcrit *epd-pgk* et de la faible efficacité de traduction de l'ORF *epd* (Boschi-Muller *et al.*, 1997; Bardey, 2000). Comme nous l'avons dit, l'ORF *fbaA* suit les ORF *epd* et *pgk* chez *E. coli*, mais aucune donnée expérimentale ne permettait de savoir si l'ORF *fbaA* est co-transcrite avec les ORF *epd* et *pgk* ou si un promoteur spécifique à *fbaA* est présent.

III.) Objectifs de mes travaux.

Une grande partie de mon travail a porté sur l'étude de la région inter-ORF *gapA-yeaA*.

Dans un premier temps, j'ai essayé de comprendre pourquoi le promoteur *gapA* P1 est extrêmement actif, alors que les deux architectures promotrices pouvant correspondre à ce promoteur n'étaient pas canoniques.

J'ai commencé par identifier laquelle de ces deux architectures possibles était fonctionnelle, puis j'ai complété une analyse structurale et fonctionnelle du complexe ouvert formé à ce promoteur que B. Charpentier avait débutée. Ceci a conduit à un article publié dans le journal *Biochemical Journal*.

Je me suis ensuite intéressé aux interrelations possibles entre la régulation du gène *gapA* et celle du gène *yeaA* dont les promoteurs transcriptionnels et éléments de régulation sont situés dans une même région. J'ai tout particulièrement étudié les rôles respectifs des différents sites potentiels de fixation de la protéine Fis, ainsi que du site de fixation du complexe CRP-AMPc et de la forte courbure intrinsèque de l'ADN sur l'expression de ces deux gènes. Ceci en étudiant l'efficacité *in vitro* et *in vivo* des promoteurs *yeaA* P1 et P2, *gapA* P1, P2 et P3 après génération de différentes mutations dans la région inter-ORF.

Etant donné que les gènes *gapA* et *pgk*, codant deux enzymes agissant en succession dans la glycolyse, sont physiquement séparés sur le chromosome d'*E. coli*, il était important d'étudier comment ils pouvaient être corégulés. B. Charpentier, V. Bardey et N. Robas avaient déjà

montré l'existence d'un mécanisme d'activation commun mais non identifié en présence de glucose dans le milieu de culture et nécessitant l'expression de la protéine EIIBC^{Glc}, transporteur spécifique du glucose (Charpentier *et al.*, 1998).

Non seulement il était intéressant d'étudier les corégulations des gènes *gapA* et *pgk* mais aussi celles de ces deux gènes avec le gène *fbaA* situé en aval de *pgk*. On ne savait pas si l'ORF *fbaA* était cotranscrite avec l'ORF *pgk*. Par ailleurs, une autre question se posait : comment l'ORF *epd* située en amont des ORF *pgk* et *fbaA* que l'on savait être, comme *pgk*, sous le contrôle du promoteur fort *gapB* P0 pouvait être très peu exprimée alors que *pgk* et *fbaA* le sont très fortement. En collaboration avec C. Vallet, j'ai donc étudié l'organisation transcriptionnelle de la région *epd-pgk-fbaA* du génome d'*E. coli* et nous avons identifié les produits de clivage obtenus après expression des transcrits. Ce travail a fait l'objet d'un article soumis à J. Bacteriol.

Enfin, le mécanisme commun d'activation des promoteurs *gapA* P1 et *gapB* P0 par le glucose nécessitant le transporteur EIIBC^{Glc} spécifique du glucose n'étant toujours pas compris, j'ai utilisé l'approche par puces à ADN pour essayer d'avoir des informations dans ce domaine. L'idée était d'identifier les gènes qui n'étaient plus exprimés ou qui étaient surexprimés en absence d'expression de la protéine EIIBC^{Glc}. Les travaux ne nous ont pas pour l'instant permis de comprendre pourquoi les promoteurs *gapA* P1 et *gapB* P0 sont activés par le glucose en présence d'EIIBC^{Glc}, mais nous avons obtenu beaucoup d'informations intéressantes sur le comportement d'*E. coli* en absence de perméation du glucose par le système PTS spécifique du glucose. Nous avons mis en évidence à la fois des changements au niveau des pores de la membrane externe et des flagelles, et ce travail fait l'objet d'un article en préparation intégré dans ce manuscrit.

Résultats
Discussion

Partie 1 : Etude du promoteur *gapA* P1 d'*E. coli*

I.) Etude de l'architecture du promoteur *gapA* P1 et du complexe ouvert (RPO) formé à ce promoteur.

I.1) Intérêt du sujet et état des travaux lorsque nous avons débuté l'étude.

Les études antérieures réalisées au laboratoire sur la GAPDH d'*E. coli* avaient mis en évidence une très forte expression de cette enzyme dans des cellules recombinantes transformées par un plasmide contenant le gène *gapA* sous contrôle de sa propre région promoteur. En effet, la quantité de GAPDH représente dans ces conditions jusqu'à 80 % des protéines solubles (Branlant *et al.*, 1983). Différents facteurs doivent intervenir dans cette forte expression. L'absence de codons rares dans l'ORF et la présence d'une séquence RBS située à une distance optimale du codon initiateur suggèrent l'existence d'une forte efficacité traductionnelle (Branlant et Branlant, 1985). Par ailleurs, la construction d'un gène hybride (séquence codant l'IL1 β humaine sous contrôle de la région promoteur du gène *gapA*) avait montré que la transcription initiée au niveau du promoteur *gapA* est très efficace (Charpentier, 1994). Cette région promoteur comporte 4 promoteurs distincts dénommés *gapA* P1 à *gapA* P4 (Charpentier, 1994). Dans un grand nombre de conditions de croissance dont celles conduisant à la forte surproduction de protéines sous contrôle de la région promoteur du gène *gapA*, le promoteur *gapA* P1 est le plus efficace de ces 4 promoteurs (Charpentier et Branlant, 1994). Il était donc tout particulièrement intéressant sur le plan biotechnologique d'étudier ce promoteur. L'examen de la séquence en amont du site P1 mettait en évidence deux couples chevauchants de boîtes -10 et -35 (P1 et P1a) (Fig. 47 A) (Branlant et Branlant, 1985).

Au cours de sa thèse, B. Charpentier avait débuté une étude du promoteur *gapA* P1 (Charpentier, 1994). Les résultats obtenus montraient que seul l'un des deux couples de boîtes (P1) est actif aussi bien *in vivo* et *in vitro*. D'après ces données, le couple de boîtes supposé constituer le promoteur, comportait une boîte -10 (TAATTT) et une boîte -35 (TTGACG) séparées par 16 pb, ce qui est inférieur à la distance optimale de 17 pb. Ce résultat était étonnant pour un promoteur très efficace. B. Charpentier avait aussi utilisé des agents chimiques, DMS (diméthyle sulfate) et KMnO₄, pour étudier la structure des complexes ouverts formés entre l'holoenzyme E σ ⁷⁰ et le promoteur *gapA* P1 *in vitro*. Le KMnO₄ modifie

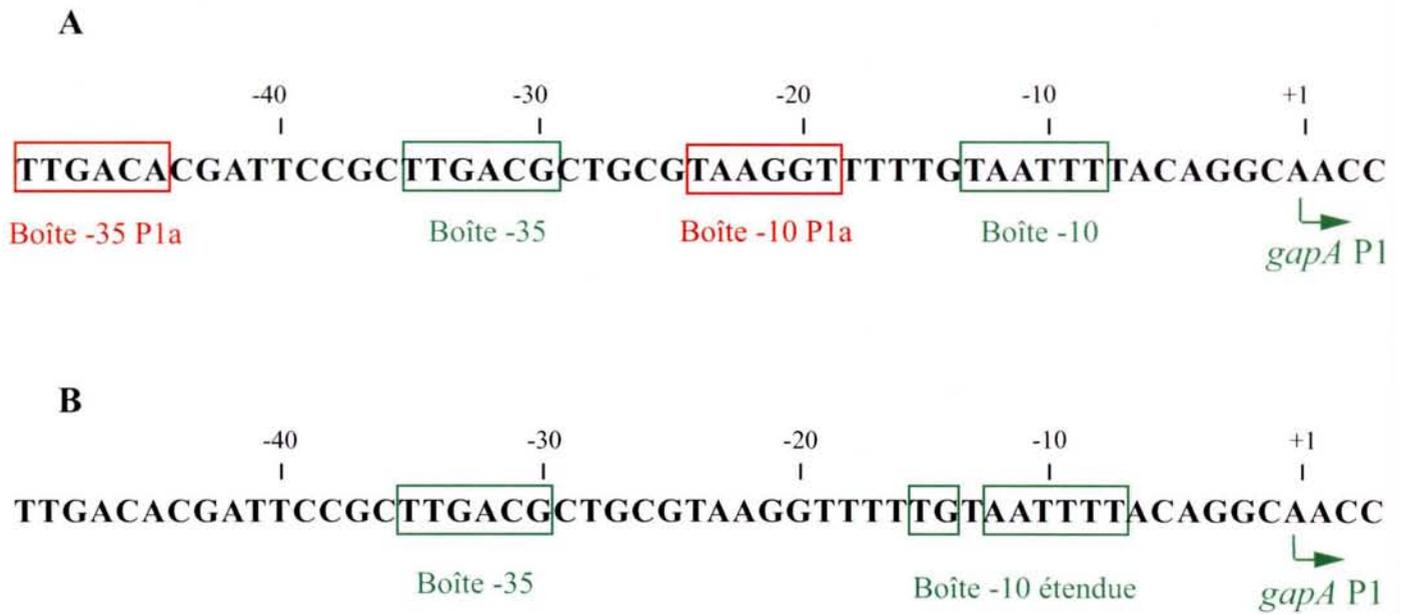


Figure 47 : Séquence du promoteur *gapA* P1 et position des boîtes -10 et -35 potentielles.

Le site d'initiation de la transcription (+1) est indiqué par une flèche.

A) Position des boîtes -10 et -35 potentielles définies par Branlant et Branlant (1985). Le couple de boîtes en vert est proposé correspondre au site d'initiation *gapA* P1 (Charpentier et Branlant, 1994)

B) Structure alternative possible du promoteur *gapA* P1 contenant une boîte -10 étendue qui s'est avérée être la structure fonctionnelle (Thouvenot *et al.*, 2004).

préférentiellement les bases T en simple brin, alors que le DMS méthyle les positions N-3A et N-7G dans les ADN en double ou simple brin et N-1A uniquement dans les ADN en simple brin (Fig. 48) (Borowiec *et al.*, 1987). La présence de torsions ou d'autres déformations dans l'ADN peut modifier la réactivité à ces positions dans les doubles brins. La présence des modifications aux positions N-1A générées par le DMS et des modifications des résidus T par le KMnO_4 bloque l'élongation d'une amorce oligonucléotidique par le fragment de Klenow de l'ADN polymérase d'*E. coli*. Les modifications en N-3A et N-7G ne bloquent pas l'élongation d'amorce par l'ADN pol. Mais un traitement de l'ADN modifié par la pipéridine permet par β -élimination de cliver la chaîne phosphodiester au niveau des résidus modifiés à ces positions. L'élongation par l'ADN pol s'arrête donc au niveau de ces sites de clivage.

Lorsqu'un complexe ouvert est traité par ces deux réactifs, le KMnO_4 permet essentiellement de mettre en évidence la partie ouverte de l'ADN, qui n'est pas masquée par l'ARN pol. L'utilisation du DMS permet de compléter cette identification de la région ouverte, mais aussi de détecter l'apparition de protections et de distorsions de l'ADN double brin en présence de l'ARN pol. En effet, l'interaction de l'holoenzyme avec l'ADN peut modifier la structure de la double hélice et ainsi la réactivité des bases vis-à-vis du DMS. De plus, l'ARN pol peut protéger des bases vis-à-vis de l'agent méthylant dans les régions où un contact étroit est établi avec l'ADN.

L'étude de la réactivité au KMnO_4 des complexes ouverts formés entre la région promoteur *gapA* P1 et l'holoenzyme $\text{E}\sigma^{70}$ purifiée mettait en évidence la formation d'un seul complexe ouvert *in vitro* correspondant au promoteur *gapA* P1 (Charpentier, 1994). Les données d'empreinte sur le promoteur *gapA* P1 n'avaient pas pu être interprétées de façon approfondie à cette époque car aucune donnée structurale n'était disponible sur le complexe RPo, ce qui n'est plus le cas aujourd'hui. Aussi, j'ai pu exploiter ces données et celles que j'ai obtenues avec l'éclairage qu'apporte maintenant la connaissance des structures 3D.

Au cours de ces études, B. Charpentier avait aussi montré la présence d'une distorsion de la structure de l'ADN au niveau du promoteur *gapA* P1, qui est provoquée par la présence de deux enchaînements de résidus T en phase avec le pas de l'hélice d'ADN. Ces enchaînements de résidus T sont localisés l'un dans la séquence de la boîte -10 et l'autre dans la région interboîte (Fig. 47). Pour cela, il avait mis au point une méthode d'étude de distorsions de l'ADN, basée sur la comparaison de la mobilité électrophorétique en gel non dénaturant de fragments de même longueur amplifiés par PCR avec des couples d'oligonucléotides conduisant à la production d'une série de fragments chevauchants décalés les uns des autres

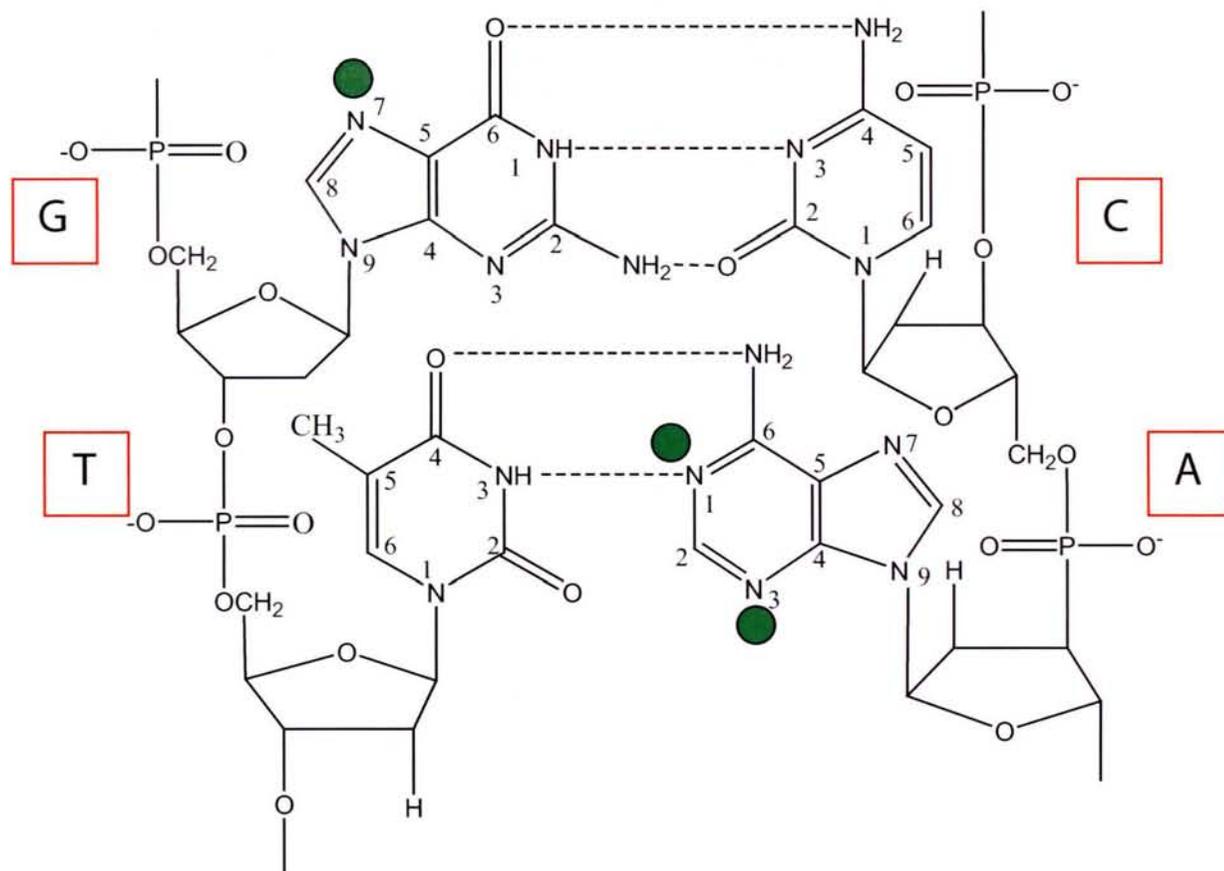


Figure 48 : Localisation des sites de modification des bases de l'ADN par le DMS (diméthyle sulfate) dans les conditions décrites par Borowiec *et al.* (1987).
 Les positions méthylées par le DMS sont indiquées par des ronds verts.

de 50 nucléotides (Charpentier, 1994). Il restait à savoir s'il s'agissait d'une courbure ou d'un autre type de déformation. Par ailleurs, l'implication possible de cette distorsion dans l'efficacité de l'initiation de la transcription au promoteur *gapA* P1 restait à démontrer.

I.2) Les objectifs de mes travaux, la démarche expérimentale et les résultats marquants obtenus.

I.2.1) Démonstration du fait que le promoteur *gapA* P1 fait partie de la famille des promoteurs étendus.

Mon premier travail au Laboratoire a consisté à approfondir l'étude des éléments de séquence du promoteur *gapA* P1. Comme nous l'avons vu, les données de la littérature montraient l'existence d'une seconde classe de promoteurs reconnus par l'holoenzyme $E\sigma^{70}$ d'*E. coli*, les promoteurs étendus. Ces promoteurs sont caractérisés par la présence d'un dinucléotide TG en amont de la boîte -10. Dans le cas du promoteur *gapA* P1, un motif TG jouxtait la boîte -10 supposée. Deux hypothèses étaient donc possibles (Fig. 47): soit le promoteur *gapA* P1 était de type « classique », avec une boîte -10 canonique (TAATTT) séparée par 16 pb de la boîte -35, soit il s'agissait d'un promoteur étendu avec une boîte -10 non canonique AATTTT. Dans cette deuxième hypothèse, la boîte -10 est séparée de la boîte -35 par la distance optimale de 17 pb.

Pour savoir laquelle de ces deux architectures possibles était celle initiant très efficacement la transcription au site P1, nous avons testé l'effet de diverses mutations générées dans les deux boîtes -10 possibles, normale et étendue, ceci dans la région promoteur du gène *gapA* P1 porté par le plasmide pBS::*EcogapA*. Ce plasmide contient aussi une partie de la séquence du gène *yeaA* transcrit sur le brin opposé (Fig. 45). L'efficacité des promoteurs mutés a été testée *in vivo* en transformant une souche d'*E. coli* par ces différentes constructions et en quantifiant les transcrits produits par ces promoteurs par la technique d'élongation d'amorce par la transcriptase inverse. L'amorce oligonucléotidique marquée était complémentaire de l'extrémité 5' de l'ORF *gapA* et donc aux ARNm initiés à partir de chacun des quatre promoteurs de ce gène. Les conditions de synthèse de l'ADNc étaient celles déjà établies au laboratoire de façon à ce que le taux d'ADNc synthétisés soit proportionnel au taux d'ARNm (Charpentier, 1994; Robas et Branlant, 1994). La position des extrémités 5' des ARNm ayant servi de matrice pour la synthèse des ADNc était identifiée par comparaison des ADNc avec les produits d'un séquençage d'ADN réalisé avec la même amorce oligonucléotidique. Il est à

noter que le taux d'ARNm estimé par cette approche résulte à la fois de l'efficacité de synthèse de l'ARNm et de leur vitesse de dégradation, il correspond à un « steady state ». Nous avons considéré que le taux de dégradation restait identique dans des conditions similaires de culture, lorsque les mutations n'affectaient que la région promoteur du gène *gapA*. Nous avons donc fait l'hypothèse que dans ce cas, l'efficacité du promoteur est directement reflétée par le taux d'ARNm présent dans l'extrait d'ARN total. Nous avons aussi vérifié que dans les conditions de synthèse d'ADNc employées, le taux de transcrits initiés à partir du gène *gapA* chromosomique était négligeable. Comme certaines des mutations prouvaient affecter le nombre de copies de plasmide, nous avons utilisé le taux de transcrits initiés à partir du promoteur *yeaA*, aussi présent dans la construction pBS::*EcogapA*, pour tenir compte des variations possibles du nombre de copies de plasmide.

Cette approche nous a permis de montrer que la séquence de la boîte -10 est la séquence non canonique AATTTT et qu'il s'agit d'un promoteur étendu, puisque le dinucléotide TG est essentiel à l'activité (Fig. 1 article n°1). La séquence de la boîte -10 est divergente par rapport à la séquence consensus des boîtes -10 des promoteurs reconnus par l'holoenzyme $E\sigma^{70}$ (cf. Fig. 1 Introduction). Nous avons donc conclu que le promoteur *gapA* P1 appartient à la famille des promoteurs étendus.

I.2.2) Démonstration du fait que le promoteur *gapA* P1 est un promoteur étendu particulier nécessitant la présence d'une boîte -35.

Pour les premiers promoteurs étendus caractérisés chez *E. coli*, la présence d'une boîte -35 ne s'était pas avérée essentielle (Bown *et al.*, 1997), ce qui peut s'expliquer par la formation d'une interaction entre le dinucléotide TG et le domaine $\sigma 3$. Cette interaction remplacerait fonctionnellement l'interaction ayant lieu entre le domaine $\sigma 4$ et la boîte -35 dans les promoteurs classiques. Néanmoins, comme nous l'avons vu dans l'introduction, les compilations faites ensuite sur l'ensemble des promoteurs d'*E. coli* (Mitchell *et al.*, 2003) montrent que certains promoteurs présentant un dinucléotide TG une paire de base en amont de la boîte -10, qui sont donc susceptibles d'être des promoteurs étendus, possèdent une séquence en amont de la boîte -10 présentant une forte similitude avec la séquence consensus des boîtes -35 des promoteurs classiques reconnus par l'holoenzyme $E\sigma^{70}$. Par contre, aucune étude de l'importance fonctionnelle possible de ces boîtes -35 n'avait été réalisée. Etant donné que le promoteur *gapA* P1 possède à la fois une boîte -10 étendue et 17 pb en amont, une

séquence fortement homologue à la séquence consensus des boîtes -35, nous avons donc un bon modèle d'étude pour tester si certains promoteurs étendus nécessitent la présence d'une boîte -35 pour qu'il y ait formation efficace du complexe RPo. En utilisant les mêmes approches que pour l'étude de la boîte -10, nous avons constaté que la substitution de la boîte -35, par une séquence complémentaire, diminue très fortement l'efficacité du promoteur *gapA* P1 (Fig. 3 article). De plus, nous avons montré par conversion de la boîte -10 du promoteur *gapA* P1 par une boîte consensus, que la nécessité d'une boîte -35 est liée au fait que la boîte -10 est non canonique. Deux explications pouvaient être imaginées, soit la boîte -35 interagit avec un facteur de régulation, soit elle interagit directement avec l'holoenzyme $E\sigma^{70}$ lors de la formation du complexe d'initiation. Pour trancher entre ces deux hypothèses, nous avons mis en œuvre une technique de transcription *in vitro* dans laquelle l'efficacité d'initiation de la transcription au promoteur *gapA* P1 était testée en présence de l'holoenzyme $E\sigma^{70}$ purifiée. Pour cela, nous avons utilisé un plasmide que nous a fourni l'équipe de R. Gourse (Université du Wisconsin), pRLG770. Ce plasmide permet d'insérer le promoteur à tester (entre les sites *EcoRI* et *HindIII*) 170 nucléotides en amont du terminateur fort de l'opéron ribosomique *rrnB* (Fig. 49 A) (Ross *et al.*, 1990), ce qui permet de produire un transcrit de 170 nucléotides à partir du promoteur dont on teste l'efficacité. Nous avons inséré le promoteur *gapA* P1 et un grand nombre de variants de ce promoteur dans ce plasmide (Fig. 49 B). Nous avons ainsi pu tester l'efficacité relative de l'initiation de la transcription de ces différents promoteurs, en présence de l'holoenzyme $E\sigma^{70}$ purifiée commerciale (Roche), ceci en comparant le taux de transcrits de 170 nucléotides produit pendant un temps donné et dans les mêmes conditions bien définies d'incubation. Pour cela, un des nucléotides utilisés pour la réaction était radiomarké. Les transcrits produits étaient fractionnés en gel de polyacrylamide dénaturant et la radioactivité était mesurée par emploi d'un PhosphorImager. Le plasmide pRLG770 code aussi de manière indépendante le transcrit RNA I (régulateur de la répliation du plasmide) dont la taille est de 108 nt. Ce transcrit nous a servi de référence pour normaliser les données obtenues pour les différents promoteurs.

I.2.3) Démonstration du fait que les interactions entre l'ARN pol et l'ADN du promoteur *gapA* P1 sont multiples.

Cette étude *in vitro* nous a permis de montrer que la boîte -35 du promoteur *gapA* P1 est requise pour la formation d'un complexe d'initiation de la transcription actif en absence de

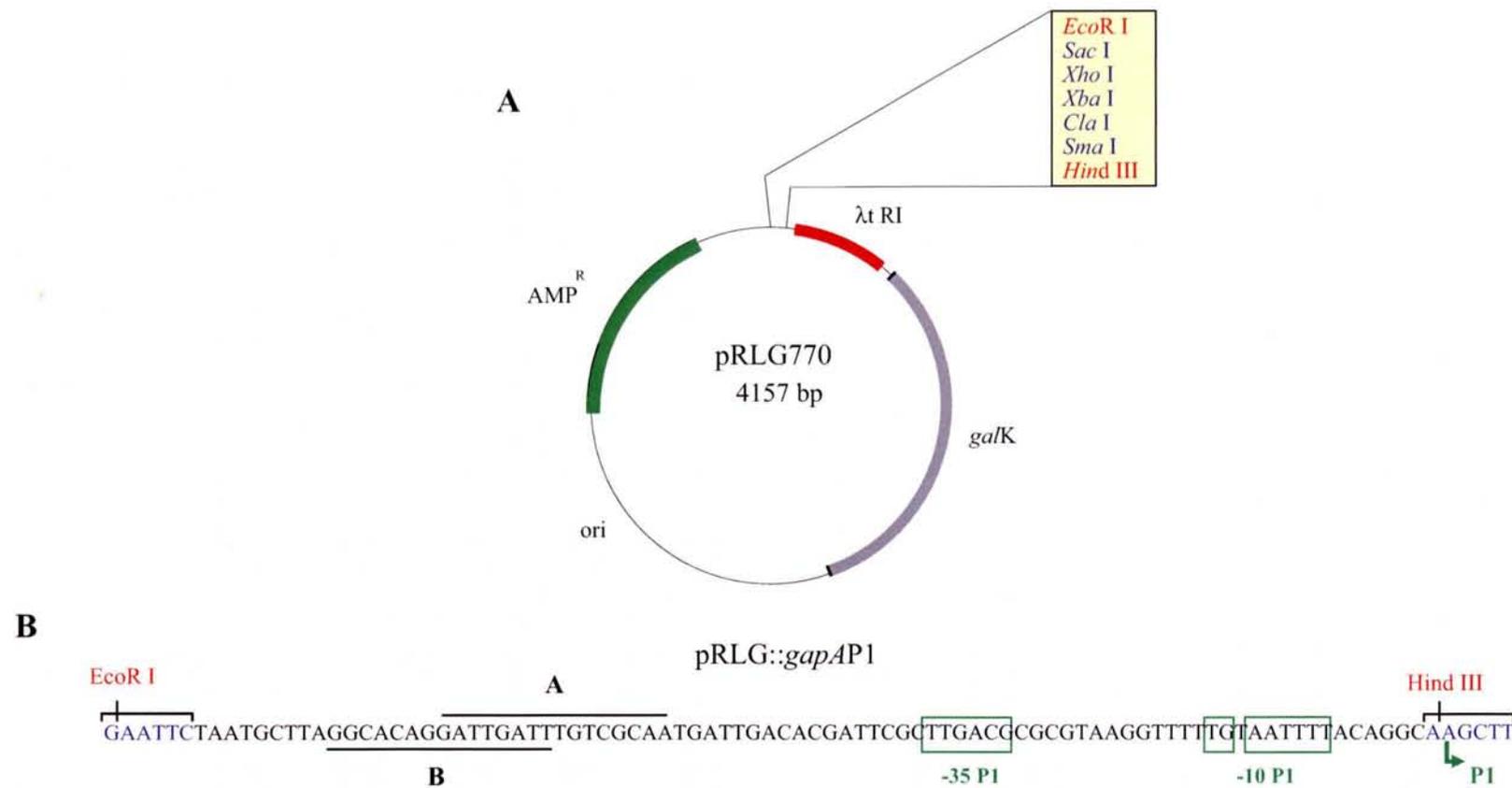


Figure 49 : Constructions réalisées pour les études de l'effet de mutations dans le promoteur *gapA* P1 sur son efficacité d'initiation de la transcription *in vitro*.

A) Carte de restriction du plasmide pRLG770 contenant les sites *EcoR* I et *Hind* III utilisés pour le clonage du promoteur P1 du gène *gapA*. Le terminateur λ t RI de l'opéron ribosomique *rrnB* permet une terminaison de la transcription initiée au niveau du promoteur inséré.

B) Région du promoteur *gapA* P1 (positions -84 à +1) insérée dans le plasmide pRLG770 par l'intermédiaire des sites de restriction *EcoR* I et *Hind* III. Le site d'initiation de la transcription est représenté par une flèche et les boîtes -10 et -35, ainsi que le dinucléotide TG sont encadrés en vert. La position des sites potentiels de fixation de la protéine Fis (A et B) est indiquée par des traits noirs.

facteur *trans* cellulaire. On pouvait donc supposer que cette boîte établissait un contact direct avec l'holoenzyme. Comme nous l'avons déjà dit, il avait été proposé que pour les promoteurs étendus, le contact établi entre le dinucléotide TG et le domaine σ_3 du facteur σ^{70} remplaçait l'interaction entre la boîte -35 et le domaine σ_4 . On pouvait donc se demander quels types de contacts étaient mis en place dans le cas d'un promoteur tel que *gapA* P1 renfermant à la fois une boîte -35 fonctionnelle et une boîte -10 étendue. Etant donné que comme nous l'avons expliqué ci-dessus, B. Charpentier avait réalisé une étude par empreinte chimique du complexe RPo formé avec le promoteur *gapA* P1 et que d'autre part un modèle de structure 3D de l'holoenzyme de *T. aquaticus* en association avec un ADN mimant en partie un complexe ouvert avait été établi (cf. Introduction) (Murakami *et al.*, 2002b), j'ai utilisé ces deux ensembles d'informations pour essayer de comprendre comment l'holoenzyme $E\sigma^{70}$ interagit avec le promoteur *gapA* P1. Comme le modèle établi pour l'holoenzyme de *T. aquaticus* n'incluait pas l'interaction entre la boîte -35 et le domaine σ_4 , j'ai aussi utilisé la structure 3D du complexe entre la boîte -35 et σ_4 établie par Campbell *et al.* (2002).

A partir de l'ensemble de ces données, nous avons pu montrer que dans le complexe RPo *gapA* P1, la bulle de transcription s'étend des positions -12 à +3 par rapport au site +1 d'initiation de la transcription. Cette ouverture complète de la séquence de la boîte -10 est assez surprenante. En effet, les données obtenues par plusieurs techniques suggèrent que la paire de bases en position -12 reste en double brin au cours du mécanisme d'initiation de la transcription (Helmann et deHaseth, 1999; Murakami *et al.*, 2002b; Tsujikawa *et al.*, 2002). Soit ces interprétations n'étaient pas correctes soit le promoteur *gapA* P1 se comporte différemment des promoteurs classiques au niveau de l'ouverture de l'ADN. Nous avons aussi pu identifier les régions accessibles et les régions protégées par l'ARN pol au niveau des régions simple brin. De plus, nous avons constaté l'existence d'une protection du dinucléotide TG par l'ARN pol, ce qui confirmait son implication dans la formation du complexe RPo. Nous détectons aussi des modifications importantes de la réactivité des résidus de la boîte -35 et de résidus en amont, ce qui est en accord avec son implication fonctionnelle. Enfin, deux résidus situés très en amont (positions -56 et -58) étaient protégés, ce qui suggère l'existence de contacts entre l'ARN pol et cette région distale. Ceci pourrait être compatible avec l'existence d'une interaction avec la sous-unité α -CTD de l'ARN pol comme cela a été montré pour certains promoteurs classiques.

I.2.4) Démonstration du rôle d'une courbure naturelle induite par deux séries de résidus T présentes dans la région interboîte et la boîte -10.

L'ensemble de ces données montrait donc une grande complexité des interactions mises en place entre l'ADN du promoteur *gapA* P1 et σ^{70} dans le complexe RPo. Par ailleurs, d'après les empreintes chimiques, des modifications de conformation semblaient avoir lieu dans la région interboîte. Nous avons rapproché cette information des deux types de données suivantes : dans le modèle établi par Murakami *et al.* (2002b), l'ADN assume des courbures au niveau de la région interboîte. De plus, B. Charpentier avait mis en évidence l'existence d'une distorsion au niveau de la région interboîte du promoteur *gapA* P1. Il était donc possible qu'une courbure naturelle existe dans le promoteur *gapA* P1 et qu'elle soit modifiée ou renforcée lors de la formation du complexe ouvert. J'ai d'abord confirmé que la distorsion était bien une courbure en utilisant la technique mise au point par B. Charpentier et décrite précédemment, sur un grand nombre de variants portant des mutations dans l'enchaînement de résidus T, présent dans la région interboîte. Par ailleurs, j'ai utilisé un logiciel disponible sur internet au site http://hydra.icgeb.trieste.it/~kristian/dna/model_it.html qui permet à partir de la séquence nucléotidique d'un fragment d'ADN de faire une prédiction de la conformation de la double hélice, en particulier, de l'existence de courbures. En parallèle, j'ai testé l'efficacité des promoteurs portant les mutations dans la région interboîte à la fois *in vivo* et *in vitro*, ce qui a permis de montrer que la présence de la courbure naturelle située dans la région interboîte augmente l'efficacité du promoteur *gapA* P1 et ceci sans qu'il y ait intervention de facteurs *trans* cellulaires.

I.2.5) Démonstration du fait que les mutations créées influencent le taux de complexe RPo.

Nous avons montré que des mutations réalisées aussi bien dans la boîte -10 que dans la boîte -35 ou la région interboîte pouvaient fortement affecter le taux de transcription *in vitro* et *in vivo*. Il était alors intéressant de préciser l'étape de l'initiation de la transcription pour laquelle ces éléments architecturaux sont importants. Pour cela, nous avons utilisé le KMnO_4 qui permet de modifier les bases T de la région ouverte de l'ADN dans le complexe RPo. Le taux de modification reflète le taux de RPo formés. Nous pouvions ainsi tester l'influence des mutations sur le taux de complexes RPo formés dans des conditions définies. Nous avons ainsi montré que les mutations du promoteur *gapA* P1 qui limitent fortement son efficacité,

limitent toutes le taux de complexes ouverts RPo. Ainsi, soit la cinétique de formation du complexe est ralentie soit le complexe formé est instable.

Ces différentes techniques ont permis de montrer que le promoteur *gapA* P1 est un promoteur étendu qui présente une organisation encore jamais démontrée pour un promoteur d'*E. coli*. En effet, la boîte -35 est nécessaire à l'efficacité et contacte l'ARN pol, de même que le dinucléotide TG de la boîte -10 étendue, qui elle est non canonique. En plus, une courbure de l'ADN est centrée à l'extrémité de la région interboîte et cette courbure est un déterminant important pour l'efficacité de ce promoteur. Enfin, un contact distal $E\sigma^{70}$ /ADN est mis en place peut-être par l'intermédiaire de la sous-unité α de l'ARN pol.

II.) Article 1 : The strong efficiency of the *Escherichia coli gapA* P1 promoter depends on a complex combination of functional determinants.

B. Thouvenot, B. Charpentier et C. Branlant, Biochem. J., 383 (2), 371-382.

Publication n°1

The Strong efficiency of the *Escherichia coli gapA* P1 promoter depends on a complex combination of functional determinants.

Thouvenot B., Charpentier B. et Branlant C.

Biochem. J., 383 (2), 371-382.

The strong efficiency of the *Escherichia coli* *gapA* P1 promoter depends on a complex combination of functional determinants

Benoit THOUVENOT, Bruno CHARPENTIER¹ and Christiane BRANLANT

Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Boulevard des Aiguillettes, Faculté des Sciences et Techniques, BP 239, 54506 Vandoeuvre-les-Nancy cedex, France

The *Escherichia coli* multi-promoter region of the *gapA* gene ensures a high level of GAPDH (glyceraldehyde-3-phosphate dehydrogenase) production under various growth conditions. In the exponential phase of growth, *gapA* mRNAs are mainly initiated at the highly efficient *gapA* P1 promoter. In the present study, by using site-directed mutagenesis and chemical probing of the RPo (open complex) formed by $E\sigma^{70}$ (holoenzyme associated with σ^{70}) RNAP (RNA polymerase) at promoter *gapA* P1, we show that this promoter is an extended -10 promoter that needs a -35 sequence for activity. The -35 sequence compensates for the presence of a suboptimal -10 hexamer. A tract of thymine residues in the spacer region, which is responsible for a DNA distortion, is also required for efficient activity. We present the first chemical probing of an RPo formed at a promoter needing both a -10 extension and a -35 sequence. It reveals a complex array of RNAP–DNA interactions. In agreement with the fact that residue A-11 in the non-template strand is flipped out in a protein

pocket in previously studied RPos, the corresponding A residue in *gapA* P1 promoter is protected in RPo and is essential for activity. However, in contrast with some of the previous findings on RPos formed at other promoters, the -12 A:T pair is opened. Strong contacts with RNAP occur both with the -35 sequence and the TG extension, so that the $\sigma 4$ and $\sigma 2$ domains may simultaneously contact the promoter DNA. RNAP–DNA interactions were also detected immediately downstream of the -35 hexamer and in a more distal upstream segment, reflecting a wrapping of RNAP by the core and upstream promoter DNA. Altogether, the data reveal that promoter *gapA* P1 is a very efficient promoter sharing common properties with both extended -10 and non-extended -10 promoters.

Key words: DNA distortion, DNA–RNA polymerase interaction, *Escherichia coli*, extended -10 promoter, *gapA* gene, open complex chemical probing.

INTRODUCTION

GAPDH (D-glyceraldehyde-3-phosphate dehydrogenase) with phosphorylating activity is a key enzyme for glucose metabolism. It catalyses the reversible oxidation of D-glyceraldehyde-3-phosphate to 1,3-diphosphoglycerate. In many bacteria species, the gene encoding the phosphorylating GAPDH (*gap* gene) precedes the *pgk* gene encoding the PGK (phosphoglycerate kinase) [1,2]. However, these genes have a different organization in γ -proteobacteria compared with other bacteria. The best-studied example is *Escherichia coli*, where an *epd* ORF (open reading frame) is located upstream of the *pgk* ORF. It encodes a protein showing homology with GAPDH, but having a non-phosphorylating erythrose-4-phosphate dehydrogenase activity [3,4]. The *E. coli* phosphorylating GAPDH is encoded by the *gapA* gene that is located at another locus on the chromosome [5,6]. Nevertheless, GAPDH and PGK are co-ordinately expressed in *E. coli*.

In an initial study [7], we showed that *gapA* transcription is governed by a multi-promoter region that responds to several of the general regulatory mechanisms used by *E. coli*. Promoters P1, P3 and P4 are recognized by the general $E\sigma^{70}$ (holoenzyme associated with σ^{70}) RNAP (RNA polymerase), whereas promoter P2 is recognized by the $E\sigma^{32}$ RNAP. Promoter P3 is subjected to catabolic repression. An interplay between the different *gapA* promoters ensures the expression of GAPDH under various environmental conditions. However, of the four *gapA* promoters,

promoter P1 plays an essential role, since transcripts initiated at promoter P1 are the most abundant ones throughout the growth period and at the beginning of the stationary phase [7]. The strong activity of promoter P1 explains the very high level of expression of the cloned *E. coli gapA* gene (80% of soluble proteins) [5]. Promoter *gapA* P1 probably also plays an important role in the co-ordinated expression of the *gapA* and *pgk* ORFs. Indeed, GAPDH and PGK production is increased in the presence of glucose in the medium and these increases are explained by an augmentation of the amounts of transcripts initiated at the *gapA* P1 promoter and *gapB* P0 promoter, which are present upstream of the *epd* and *pgk* ORF tandem [8]. Such increases in GAPDH and PGK production are not observed in a *ptsG*[−] strain, suggesting that the amounts of transcripts initiated at both the *gapA* P1 and *gapB* P0 promoters depend on the uptake of glucose through its specific transporter, EIIBC^{Glc}, encoded by the *ptsG* gene [8]. In addition, promoter P1 is also very efficiently used *in vitro* by the $E\sigma^{70}$ RNAP, which is an indication of strong intrinsic activity [7]. On the basis of all these observations, an in-depth study of the *gapA* P1 promoter was of high interest.

Despite having a high intrinsic efficiency [5], the promoter structure is not well defined. Indeed, the *gapA* P1 promoter has a clear -35 element (TTGACG) and two possible -10 sequences: either a canonical -10 sequence spaced 16 bp from the -35 hexamer or an extended -10 sequence with a non-canonical -10 hexamer, which is spaced 17 bp from the -35 sequence. When they were first discovered, the extended -10 promoters

Abbreviations used: α -CTD, C-terminal domain of the α subunit; DMS, dimethylsulphate; DTT, dithiothreitol; $E\sigma^{70}/E\sigma^{32}$, holoenzyme associated with σ^{70} or σ^{32} ; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; ORF, open reading frame; PGK, phosphoglycerate kinase; RNAP, RNA polymerase; RPo, open complex; *Taq*, *Thermus aquaticus*; WT, wild-type.

¹ To whom correspondence should be addressed (email bruno.charpentier@maem.uhp-nancy.fr).

of *E. coli* were considered to be constitutive promoters [9]. Indeed, the TG dinucleotide at positions -14 and -15 is recognized by domain 3.0 (2.5) of factor σ^{70} [9–13] and this interaction can compensate for the absence of a -35 sequence [14–18]. However, a recent, in-depth inspection of *E. coli* promoters revealed that TG-containing promoters are more frequent than previously expected (20% of the promoters) [19,20] and that 74% of them have a -35 hexamer with 3/6 or more matches to the -35 consensus sequence [20]. The functional importance of the TG dinucleotide upstream of the -10 hexamer was experimentally demonstrated for seven of them [20]. However, the requirement of the -35 element was not tested in parallel. The optimal length of the spacer region in *E. coli* promoters recognized by the $E\sigma^{70}$ RNAP is 17 bp, allowing interactions of the -10 and -35 hexamers with two distinct domains of factor σ^{70} (regions 2.4 and 4.2 respectively) [21–25] (see [26,27] for reviews). On the basis of this knowledge, an experimental analysis of the *E. coli gapA* P1 promoter was needed to identify its functional sequences.

In the present study, by using site-directed mutagenesis, we demonstrate that promoter *gapA* P1 belongs to the class of extended -10 promoters and needs both a TG dinucleotide and a -35 sequence for activity. We also show the importance of the sequence of the spacer region for promoter conformation and activity. Chemical probing of the RPo (open complex) structure revealed a complex array of DNA-protein interactions. This is the first complete chemical probing of a natural promoter from the extended -10 class that needs a -35 hexamer for activity. The results we obtained are discussed taking into account the three-dimensional model of the complex formed by the *Taq* (*Thermus aquaticus*) RNAP and a fork-junction promoter DNA containing an extended -10 sequence (RF complex) [12].

EXPERIMENTAL

Bacterial strains and growth conditions

Propagation of recombinant BS_{II}SK⁺ phagemids (Stratagene, La Jolla, CA, U.S.A.) was performed in *E. coli* DH5 α . Epicurian ColiTM XL1-Blue cells were used to study the *in vivo* GAPDH production and the transcriptional activity of the *gapA* P1 variant promoters. Bacteria were grown at 37 °C in LB (Luria-Bertani) broth medium, and 100 mg/l ampicillin was added for the growth of transformed cells.

Genetic constructions and site-directed mutagenesis

Plasmid pBS::*EcogapA* [7] was used as a reference for GAPDH and mRNA production by the WT (wild-type) *gapA* gene. In this construct, the *gapA* ORF is preceded by a 483 bp upstream non-coding sequence containing the multi-promoter region. To create variants (mt-9, -TG, -12, -17T to C, -19T to G, -19T to C, -19T to G, -19T to A, -35, -sp3'), site-directed mutageneses were performed on plasmid pBS::*EcogapA*. Other variants (mt-16T to G, -16T to C, -18T to C, -18T to G, -18T to A, -cons-10, -cons-10/mt-35, mt-20G to C, mt-20G to T) were obtained by site-directed mutagenesis of plasmid pBS::*EcogapAmut62* with a Quik ChangeTM Site-Directed Mutagenesis kit (Stratagene). This plasmid carries the same promoter region as plasmid pBS::*EcogapA*, except that an *NdeI* restriction site was introduced at the ATG codon. The presence of this restriction site does not change the *in vivo* levels of *gapA* mRNAs and GAPDH when compared with the WT sequence (results not shown). After mutagenesis, a *HindIII*-*NdeI* DNA fragment carrying the mutated sequence was cloned between the *HindIII* and *NdeI* sites of plasmid pBS::*EcogapAmut62*.

Derivatives of the pRLG770 plasmid [28] were built for *in vitro* run-off transcription assays. To this end, DNA fragments containing the WT or variant *gapA* P1 promoters were PCR-amplified from position +1 to -84 (+1 corresponds to the transcription-start site) using two primers (1280 and 1278), which introduced an *EcoRI* site at the 5'-end and a *HindIII* site at the 3'-end respectively. To obtain the series of pRLG::*gapA*P1 plasmids, the *EcoRI*-*HindIII* fragments were ligated into the pRLG770 plasmid [28] that was digested by these two enzymes.

Protein extract preparation and enzymic assay

Enzymic activities were measured on the soluble fraction obtained after sonication of the cells harvested at an absorbance A_{600} 0.5. GAPDH activity was measured as described previously [8]. The β -lactamase activity measurements were performed as described by Sargent [29].

Extraction of total RNA and mRNA analysis

Total RNA was isolated as described previously by the hot phenol extraction procedure [7]. RNA extractions were performed on a fixed number of cells (5×10^8) grown at 37 °C in LB/ampicillin medium until the A_{600} reached 0.5. The *gapA* mRNAs, as well as the truncated *yeaA* mRNAs also encoded by plasmid pBS::*EcogapA*, were analysed by primer extension, using conditions described previously that allow quantitative measurement of mRNA levels [7]. Briefly, 5 μ g of total RNA isolated from transformed cells was annealed with 5 ng of 5'-end-labelled *gapA* primer 1043 (5'-TTTCTGAGCAGCACGGAA-3') or *yeaA* primer 437 (5'-GATTCTGCGTCACGT-3') at 65 °C for 10 min. Primer extensions were performed for 30 min at 45 °C with 1 unit of AMV (avian myeloblastosis virus) Reverse Transcriptase (Life Science) in 50 mM Tris/HCl (pH 8.3), 40 mM KCl and 6 mM MgCl₂. The synthesized cDNAs were fractionated by electrophoresis on a 6% sequencing gel and the fractionation pattern was revealed by autoradiography. Since the WT and variant *gapA* mRNAs initiated at the P1 promoter had identical lengths and sequences, they were expected to have identical half-lives under similar growth conditions. Hence, we assumed that the steady-state levels of *gapA* P1 mRNA in cells carrying the variant genes were directly correlated to *gapA* P1 promoter efficiency. As a prerequisite for a quantitative measurement of the *gapA* P1 mRNA level, we verified that the contribution of the chromosomal *gapA* gene to *gapA* P1 mRNA level was negligible. To take into account possible variations of the plasmid copy number in the different cultures, the steady-state level of the *yeaA* transcript was used as an internal reference. The part of the *yeaA* gene that is present in plasmid pBS::*EcogapA* directs transcription initiation on the opposite strand compared with promoter *gapA* P1 (B. Charpentier, unpublished work). The ratio between *gapA* P1 and *yeaA* mRNA levels was estimated for each experiment by measuring the radioactivity contained in the cDNA bands using a PhosphorImager. The variations of *gapA* P1 mRNA levels in cells containing the variant genes were expressed as a percentage of the *gapA* P1/*yeaA* P1 mRNA level found in cells transformed with the WT pBS::*EcogapA* plasmid. For each variant *gapA* P1 gene, the experiment was repeated three times on independent cell cultures and the values given are the mean values for the three experiments. The range error on this mean value was estimated using Excel software.

In vitro run-off transcription

Plasmid pRLG::*gapA*P1 and its derivatives were used as supercoiled DNA templates for *in vitro* run-off transcription assays.

Plasmids (0.8–1 nM) were incubated for 15 min at 22 °C in 25 µl of 10 mM Tris/HCl (pH 8), 150 mM NaCl, 10 mM MgCl₂, 1 mM DTT (dithiothreitol) and 100 µg/ml BSA. Then, heparin (10 µg/ml) was added and incubation was continued for 5 min before the addition of a mixture of 500 µM ATP, 200 µM CTP and GTP, 10 µM UTP, 5 µCi of [α -³²P]UTP and 0.2 unit of RNAP (E σ ⁷⁰; Roche). The reaction was stopped after 20 min by the addition of 5 µl of a stop solution [10 mM EDTA, 1% SDS, 2 × TBE (180 mM Tris/borate, pH 8.3, and 4 mM EDTA) and 8 M urea, 0.05% Bromophenol and 0.025% xylene cyanol]. Samples were loaded on to 5% polyacrylamide denaturing gels and the radioactivities contained in the DNA bands corresponding to the *gapA* P1-*rrnB* and RNAI transcripts were measured using a PhosphorImager with the Image Quant software (Molecular Dynamics). Variations of the levels of the *gapA* P1 transcript in assays performed with the variant promoters were expressed as a percentage of the ratio of *gapA* P1 transcript/RNAI transcript found for the WT promoter. The experiment was repeated three times for each variant promoter and the values given are the mean values for three independent experiments. The range error on the mean value for three different experiments was estimated using Excel software.

Analysis of DNA conformation

The following method was used for the PCR amplification of DNA fragments carrying the WT or variant *gapA* P1 promoters. The supercoiled plasmid pBS::*EcogapA* (50 ng) was denatured for 3 min at 94 °C in 40 µl of 10 mM Tris/HCl (pH 9), 50 mM KCl, 1.5 mM MgCl₂, 0.01% (w/v) gelatin and 0.1% Triton X-100, in the presence of 100 ng of each primer 110 and 351 (5'-TATTTTCAGCACATATGCCATG-3' and 5'-ATTGCCCTTTAAAATTTCGGG-3' respectively). Then, 0.2 mM of each dNTP and 0.25 unit of *Taq* DNA polymerase were added to a final volume of 50 µl. A series of 30 cycles was performed in a PerkinElmer Thermal Cycler with incubations at 94 °C for 30 s for denaturation, 55 °C for 30 s for hybridization and 72 °C for 1 min for polymerization. The amplified DNAs were extracted with a phenol/chloroform/isoamylalcohol (24:24:1, by vol.) mixture, precipitated with ethanol, dissolved in TE buffer (1 mM EDTA and 1 mM Tris/HCl, pH 8.0) and fractionated by electrophoresis on a 6% polyacrylamide non-denaturing gel. A 29:1 ratio of acrylamide/bisacrylamide was used. Electrophoresis was performed at 4 °C in Tris/borate buffer (2 mM EDTA and 90 mM Tris/boric acid, pH 8.5) at 7 V/cm. The *Sau3AI* digestion products of pBR322 were used as molecular-mass standards. After electrophoresis, the gels were stained with ethidium bromide (0.5 µg/ml) and DNA was visualized by UV illumination.

The DNA bending model of WT and variant *gapA* P1 promoters were made by using the plot.it server (<http://www.icgeb.trieste.it/dna/>) [30]. WebLab Viewer Pro was used to visualize and represent the different models.

Chemical probing of DNA

The reactivity of promoter DNA towards DMS (dimethylsulphate) or potassium permanganate (KMnO₄) in the absence or presence of purified E σ ⁷⁰ RNAP was studied by the method of Borowiec et al. [31]. A supercoiled template (5 nM) was incubated for 30 min at 37 °C with the purified commercial E σ ⁷⁰ holoenzyme (100 nM) in 50 µl of buffer I [40 mM Tris/HCl (pH 8), 30 mM KCl, 10 mM MgCl₂, 1 mM DTT and 300 µg/ml BSA] for plasmid pBS::*EcogapA* or in 50 µl of buffer II [10 mM Tris/HCl (pH 8), 150 mM NaCl, 10 mM MgCl₂ and 100 µg/ml BSA] for plasmid pRLG*gapA*P1. Heparin was added to a final concentration

of 10 µg/ml before adding the chemical reagent. For DMS modification, DMS was added to a 5 mM concentration. After 5 min of incubation at 37 °C, the reaction was stopped by the addition of 12.5 µl of a G-stop solution [1.5 M sodium acetate (pH 7) and 1 M 2-mercaptoethanol], and the DNA was ethanol-precipitated, washed with 70% (v/v) ethanol, dried and dissolved in 100 µl of 1 M piperidine. After a 30 min incubation at 90 °C, the DNA was again precipitated with ethanol. For KMnO₄ modification, KMnO₄ (8 mM) was added for 2 min. The reaction was stopped by the addition of 1.2 M 2-mercaptoethanol followed by ethanol precipitation. The pellets from DMS and KMnO₄ modifications were washed with 70% ethanol, dried and dissolved in 35 µl of MilliQ water.

Positions modified by KMnO₄ and DMS were identified by primer extension by the method of Borowiec et al. [31]. Alkaline denaturation of the double-strand plasmid was performed with NaOH (1 mM) and 5 ng of 5'-end-labelled oligonucleotide [1043 (sequence shown above) or 352 (5'-GGAAGAGTGAGGCG-3') for plasmid pBS::*EcogapA* or 2559 (5'-ACCATCGGCGCTACGGCG-3') for plasmid pRLG*gapA*P1] were added. The DNA was denatured at 80 °C for 2 min and cooled on ice for 5 min. Samples were then neutralized with 5 µl of 10 × buffer [500 mM Tris/HCl (pH 7.2), 100 mM MgSO₄ and 2 mM DTT] and heated for 3 min at 45 °C for annealing. Samples were cooled again on ice for 5 min before the addition of 5 µl of a dNTP mixture (5 mM each). Primer extensions were performed for 10 min at 50 °C with 1 unit of Klenow fragment enzyme. The reaction products were precipitated, and the pellet was dissolved in formamide dye buffer and analysed by electrophoresis on 7% sequencing gels. Dideoxy-sequencing reaction products, obtained with the same primers, were run in parallel.

RESULTS

Activity of the *gapA* P1 promoter depends on an extended – 10 hexamer

The two possible architectures (P1 and P1) of promoter *gapA* P1 are shown in Figure 1(a). To test for the activity of each putative – 10 element, their 5'-terminal dinucleotide and 3'-terminal residues, which are known to play an important role in the initiation process [12,20,22], were substituted for a GG dinucleotide (mt-2 to test for the P1 – 10 sequence and mt-9 to test for the P1 – 10 sequence) and a G residue (mt-5 for the P1 – 10 sequence and mt-12 for the P1 – 10 sequence) respectively (Figure 1a). For an *in vivo* analysis of their activities, the mutated promoters were cloned into plasmid pBS::*EcogapA* [7]. The amounts of *gapA* transcripts were estimated by primer-extension analysis with reverse transcriptase under the conditions described in the Experimental section (Figures 1a and 1b), using an oligonucleotide primer complementary to the *gapA* ORF. To take into account possible variations of the plasmid copy number, the steady-state level of transcripts from the 5' part of the *yeaA* gene, which is located upstream of the *gapA* gene and is transcribed from the opposite strand (B. Charpentier, unpublished work), was used as an internal reference. GAPDH production in recombinant cells was also monitored by enzymic activity measurement of cell extracts, as described in the Experimental section (Figure 1c). The amount of mRNAs initiated at the *gapA* P1 promoter was significantly reduced by mutations at the 5' or 3' extremities of the P1 – 10 sequence (AATTTT). In contrast, no decrease was observed when the TA dinucleotide at the 5'-end of the P1 – 10 hexamer (TAATTT) was mutated (Figure 1c, cf. lanes mt-9 and mt-12 with lanes mt-2 and mt-5). The activity was divided by a factor of 2 for the mt-5 mutation. Similar variations

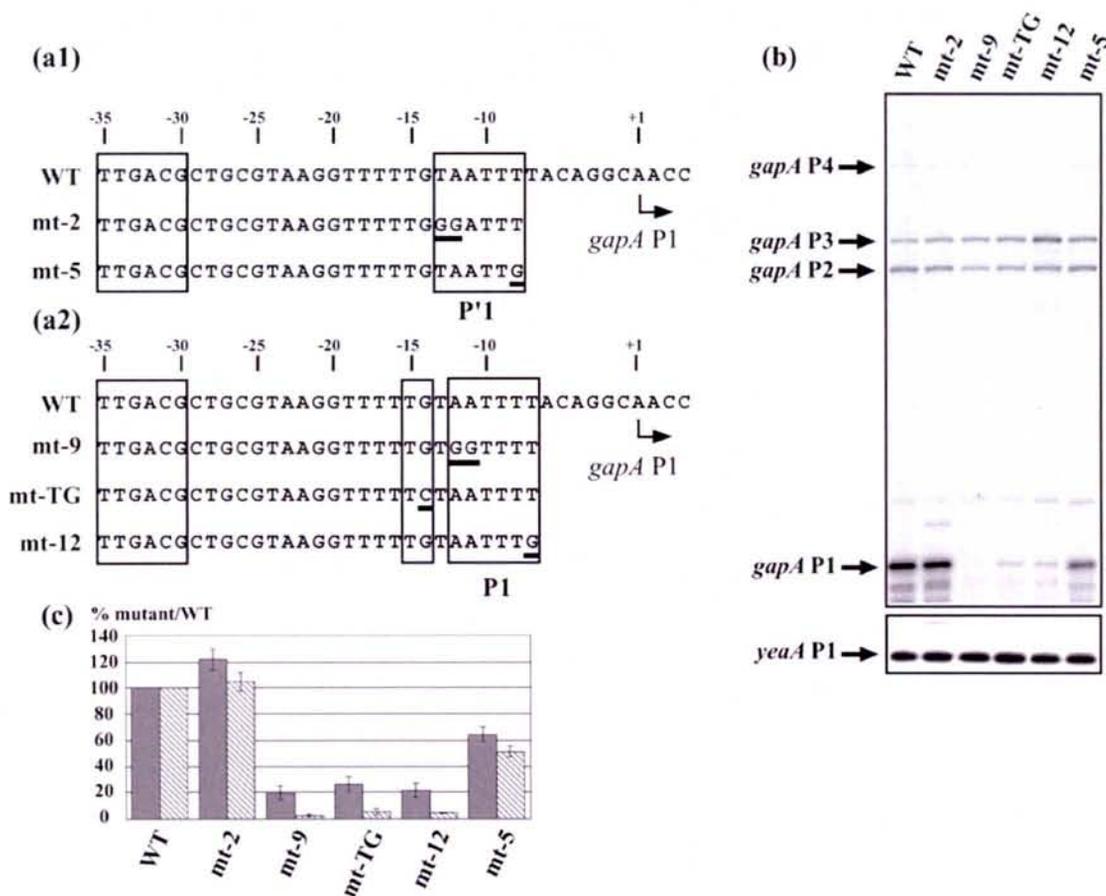


Figure 1 Promoter *gapA* P1 belongs to the class of extended -10 promoters

(a1, a2) Nucleotide sequences of the two alternative $-10/-35$ P1 **(a1)** and P1 **(a2)** promoters present in the *gapA* P1 promoter region are shown. The -10 hexamers, the TG extension and the -35 hexamer are boxed. A broken arrow marks the transcription start site (+1). Sequences of the variant *gapA* P1 promoter regions are shown below the WT sequence [variants with mutations in the P1 -10 sequence **(a1)** and variants with mutations in the P1 extended sequence **(a2)**]. Positions of the base substitutions are underlined. **(b)** Analysis of the effects of mutations on the *in vivo* *gapA* mRNA levels by reverse transcription. *Epicurian Coli*TM XL1-Blue cells transformed with plasmid pBS::*EcogapA* carrying the WT or variant *gapA* P1 promoter were grown to an absorbance A_{600} 0.5. The *gapA* mRNAs were analysed by primer extension under conditions allowing quantitative measurements (see the Experimental section). cDNAs corresponding to transcripts initiated at the P1, P2, P3 and P4 promoters are indicated. The *yeaA* mRNA primer-extension analysis (*yeaA* P1) was used for quantifying the effects of mutations on *gapA* P1 activity (see the Experimental section). **(c)** A histogram representing the effects of the mutations on GAPDH production and the *in vivo* *gapA* P1 mRNA level. GAPDH-specific activities and the plasmid-encoded β -lactamase-specific activities were measured as described in the Experimental section and their relative ratios in strains transformed with plasmids carrying the variant *gapA* genes were compared with that measured in the strain transformed with the WT pBS::*EcogapA* plasmid. In **(c)**, variations of GAPDH-specific activities were represented as percentages of the GAPDH/ β -lactamase ratio found for the WT plasmid (grey bars). The cDNA signals corresponding to *gapA* P1 and *yeaA* P1 mRNAs were quantified using a PhosphorImager. Variations of the *gapA* P1/*yeaA* P1 ratios (stippled bars) in strains transformed with plasmids carrying the variant genes were expressed as a percentage of the *gapA* P1/*yeaA* P1 ratio found in the strain transformed with the WT pBS::*EcogapA* plasmid. The values given in **(c)** correspond to the mean values for three different experiments and the estimated errors are indicated by vertical bars.

were reproducibly obtained in several independent series of experiments. In agreement with these observed variations of the *gapA* P1 mRNA levels, the rates of GAPDH production were decreased by factors of between 4 and 6 for the variant mt-12 and mt-9 genes. The lower decreases in GAPDH activity compared with *gapA* P1 mRNA levels are explained by the contribution of promoters P2 and P3 to GAPDH production. Taken together, these results suggested that the AATTTT sequence, in spite of its divergent sequence compared with the consensus -10 sequence, is the functional -10 hexamer sequence of promoter *gapA* P1. The functionality of the extended -10 architecture was confirmed by the strong decrease in promoter *gapA* P1 efficiency after substitution of the TG dinucleotide upstream of the AATTTT sequence for a TC dinucleotide (Figure 1b, lane mt-TG). As a result, GAPDH production was low (Figure 1c). Taken together, these results demonstrated that promoter *gapA* P1 belongs to the class of extended -10 promoters.

Structural analysis of the RPo formed at promoter *gapA* P1

For a complete demonstration that promoter *gapA* P1 was an extended -10 promoter, RPos were formed with purified $E\sigma^{70}$ RNAP and the supercoiled pBS::*EcogapA* plasmid, as described in the Experimental section. DMS was used to probe the single-stranded adenine residues and to test for guanine methylation at position N-7 in single- and double-stranded regions (Figures 2a and 2b). $KMnO_4$ was used to probe the single-stranded thymine residues (Figures 2c and 2d). Positions of the modifications were identified by primer extension analysis using the Klenow fragment of the *E. coli* DNA polymerase. In the RPos formed with promoter *gapA* P1, a large number of hypersensitive sites were detected on both strands of the -10 hexamer. This allowed us to map the DNA single-stranded region from position -12 (included) to $+3$ (included), taking position $+1$ as the transcription initiation site (Figure 2e).

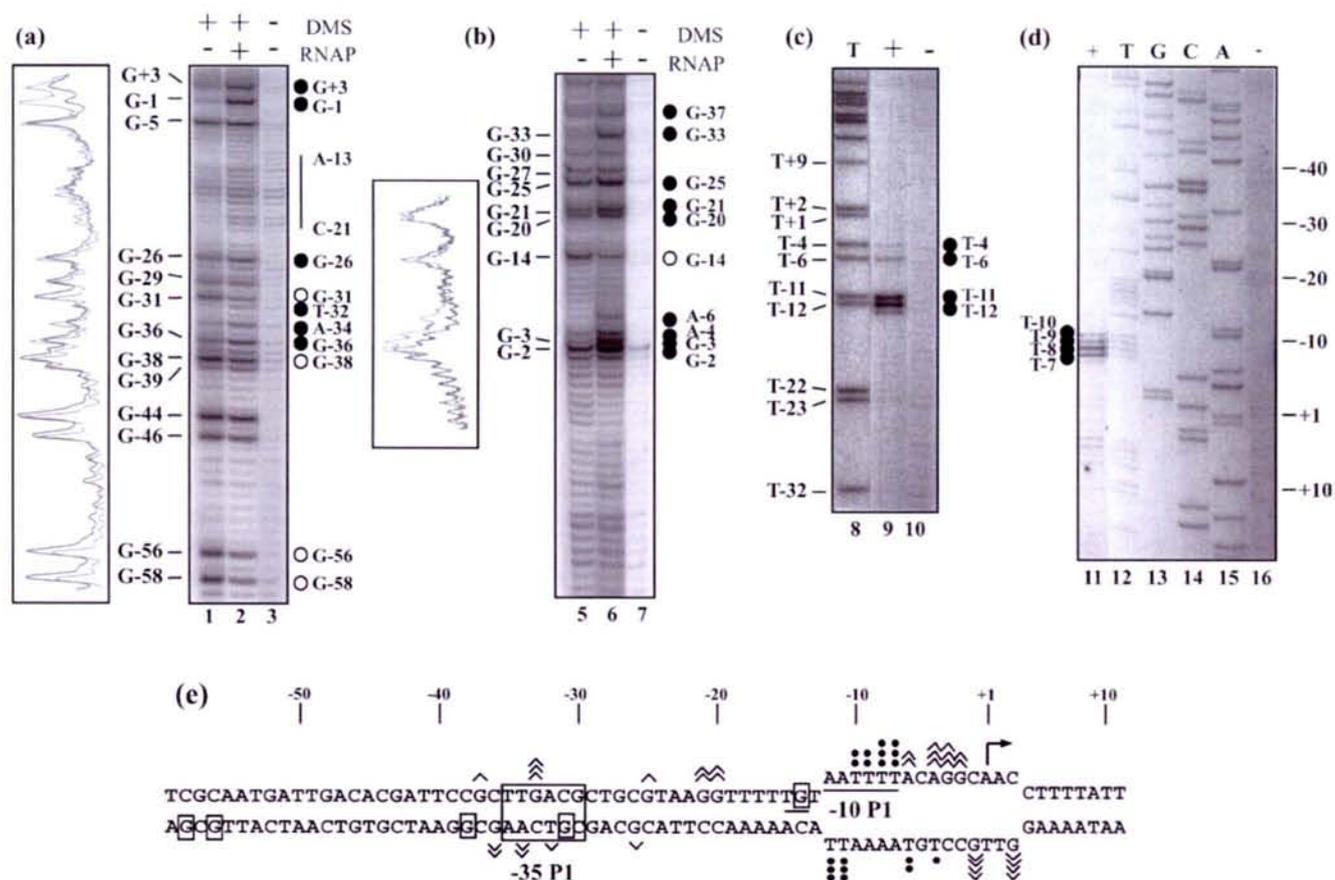


Figure 2 Numerous DNA-protein contacts take place in the *gapA* P1 RPO

RPOs were formed with the commercial $E\sigma^{70}$ RNAP and the supercoiled pBS::*EcogapA* plasmid under conditions described in the Experimental section. Probing experiments were performed on the RPO (RNAP+) and on the naked DNA (RNAP-) with DMS (a, b) or $KMnO_4$ (c, d) (see the Experimental section). DMS modifications were followed by piperidine treatment for strand cleavage at modified G and A residues. A control elongation was performed without chemical treatment [lane 3 in (a) and lane 7 in (b)]. Positions of DMS and $KMnO_4$ modifications were identified by primer extension using primer 352 for the template strands (a) and (c) and primer 1043 for the non-template strands (b) and (d). The T sequence ladder in (c) and A, C, G and T sequence ladders in (d) were obtained with primer 1043. Positions of the modified residues in the *gapA* P1 sequence are given on the left of the panels. The transcription start site was at position +1. Nucleotides that become protected in RPOs or nucleotides with an enhanced reactivity compared with DNA are indicated by open and full circles respectively. Densitograms of portions of lanes 1, 2, 5 and 6 were made with the Molecular Analysis software (Bio-Rad). Black lines correspond to lane 1 or 5 (RNAP-) and grey lines to lane 2 or 6 (RNAP+). (e) Compilation of the results of the probing experiments. Angular brackets and closed circles indicate positions of nucleotides whose reactivity to the probe (DMS or $KMnO_4$) is enhanced in the presence of RNAP. Their number is proportional to the level of increase. Boxed nucleotides were less sensitive to the chemical probe in the RPO than in the naked DNA. The -10 sequence and the TG extension are underlined and the -35 sequence is boxed. The opened DNA region is shown as well as the transcription initiation site (broken arrow).

In favour of the proposal that promoter *gapA* P1 belongs to the class of extended -10 promoters, the TG dinucleotide extension was protected against DMS methylation in the presence of RNAP (residue G-14, Figures 2b and 2e). Our probing experiments revealed strong variations in the accessibility of the putative *gapA* P1 -35 hexamer in the RPO compared with naked DNA. In the RPO, residue G-31 in the template strand was protected in the presence of RNAP (Figures 2a and 2e), whereas residues A-34 and T-32 showed an increased reactivity to DMS and, in the non-template strand, residue G-33 had a significantly increased sensitivity to DMS (Figures 2a and 2b). Moreover, upstream of the -35 hexamer sequence, residues G-36 and G-38 in the template strand had modified reactivities in the RPO (increased and decreased respectively; Figure 2a). Altogether, the strong variations of DMS modifications found within and upstream of the putative -35 hexamer strongly suggested a functional role for this hexamer, and we tested this hypothesis as described below.

Our RPO probing experiments also revealed variations of DMS methylation in the spacer region. Altogether, in the two strands

of the spacer region, four of the six G residues had an increased sensitivity to DMS in RPO (residues G-20 and G-21 and, to a lesser extent, residue G-25 in the non-template strand and residue G-26 in the opposite strand; Figures 2a and 2b). Finally, our probing data also suggested an interaction of the RNAP holoenzyme with an upstream DNA sequence of promoter *gapA* P1, since residues G-56 and G-58 were strongly protected against DMS methylation in RPO (Figure 2a). In conclusion, our probing experiments revealed a tight and complex interaction of the $E\sigma^{70}$ RNAP holoenzyme with promoter *gapA* P1.

The -35 sequence is required for promoter activity

To test for the functional importance of the putative -35 hexamer of promoter P1, we substituted the 5'-terminal TTGA sequence for an AACT sequence (mt-35; Figure 3a). No transcripts initiated at the mt-35 *gapA* P1 promoter were detected *in vivo* by primer-extension analysis (Figure 3b). Only initiation at the *gapA* P2 and P3 promoters was detected. To test whether the mutation

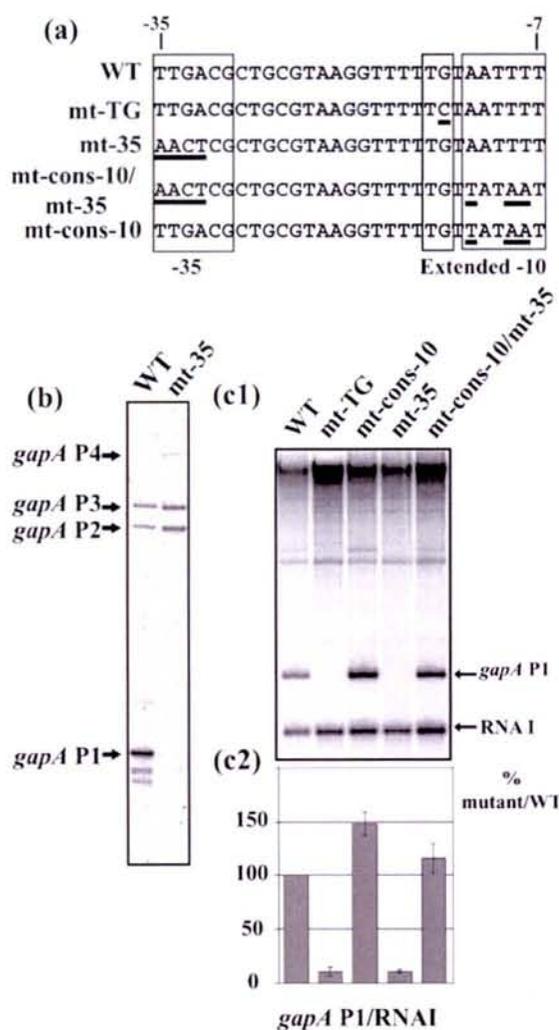


Figure 3 The -35 sequence is required for *gapA* P1 activity

(a) Sequences of the variant *gapA* P1 promoters produced to test for the importance of the -10 and -35 sequences are shown. Positions of base substitutions are underlined. (b) Effect of absence of the -35 sequence on the *in vivo* level of *gapA* mRNAs. The RNA extracts were prepared and analysed as described in the legend to Figure 1. (c1) Plasmid pRLG::*gapA*P1 carrying the WT *gapA* P1 promoter and its derivatives carrying the mutated promoters were used as templates for the *in vitro* run-off transcriptions [28]. After transcription, the products were fractionated on a 5% polyacrylamide denaturing gel as described in the Experimental section. The RNA I transcript, encoded by the plasmid, was used for the normalization of the amount of *gapA*P1 transcript recovered in each transcription assay. (c2) For each variant promoter, the *gapA* P1-*rrnB* transcript/RNAI transcript ratios were estimated using a PhosphorImager. The *gapA* P1 promoter activity of the variant promoters is represented as a percentage of the *gapA* P1-*rrnB* transcript/RNAI transcript ratio of the WT promoter. The values given in (c) are mean values for three different experiments and the estimated error is indicated by vertical bars.

influenced the binding of cellular factor(s) or had a direct negative effect on promoter P1 efficiency, *in vitro* transcription assays were performed with purified $E\sigma^{70}$ RNAP. To this end, we produced a series of pRLG::*gapA*P1 plasmid derivatives that carried the sequence coding the transcription terminator of the *E. coli* *rrnB* operon, under the control of WT or variant *gapA* P1 promoters (Figures 3c1 and 3c2). Since plasmid pRLG::*gapA*P1 derivatives also encode RNAI [28], the ratio between the amounts of *gapA* P1-*rrnB* and RNAI transcripts was used for the normalization of the data. To compare the effects of mutations in the -35 hexamer and the TG dinucleotide extension on the *in vitro* efficiency

of promoter P1, the mt-TG *gapA* P1 variant promoter was included in the experiment. We reproducibly found that alteration of the -35 sequence has the same strong negative effect as mutation of the TG dinucleotide (compare lane WT with lanes mt-TG and mt-35 in the *in vitro* transcription experiment shown in Figure 3c1). Transcripts initiated at the mt-TG and mt-35 promoters were nearly undetectable by *in vitro* transcription (Figure 3c2). We concluded that the high efficiency of the *gapA* P1 promoter depends on the presence of both a -35 hexamer and a TG extension. One possible reason could be the sequence divergence of the -10 hexamer compared with the consensus sequence (AATTTT versus TATAAT). Thus we analysed the effect of the conversion of the WT -10 sequence into a consensus -10 sequence in the presence (variant mt-cons-10) or the absence (variant mt-cons-10/mt-35) of an active -35 hexamer (Figure 3a). The *in vitro* efficiency of the variant promoter (mt-cons-10 P1), containing a -10 consensus hexamer and the WT -35 hexamer, was approx. 1.5-fold superior to that of the WT promoter (Figure 3c). In the absence of an active -35 hexamer, conversion of the WT sequence into a consensus sequence restored the WT activity (Figure 3c). Taken together, these results demonstrated that a -35 hexamer is needed for the efficient activity of the *gapA* P1 promoter because of a suboptimal -10 hexamer.

Role of the spacer sequence in promoter P1 efficiency

The appearance of additional DMS modifications in the *gapA* P1 spacer region upon RPo formation (Figure 2) suggested its distortion upon RNAP binding. DNA distortions of the spacer region were found by X-ray analysis of the complex formed between the *Taq* RNAP and a DNA promoter segment: DNA is bent around the RNAP and this is due to kinks centred at base-pair positions -35 (36°) and -25 (8°) and due to a sharp bend (37°) at base-pair position -16 [12]. Interestingly, the *gapA* P1 promoter contains a 5-T tract and a 4-T tract spaced by 8 bp (Figure 4a). Formation of intrinsic kinks and bends in double-stranded DNA is known to be sequence-dependent. In particular, the presence of multiple dAn and dTn ($n=4-6$) runs aligned in phase with the B-DNA helix repeat are known to induce DNA bends [32]. We tested whether the 5-T tract, which is located in the spacer region, may influence promoter activity. To this end, each of the 4-T residues located from positions -16 to -19 were substituted by C and G residues (Figure 4a). The effect of a T to A substitution was tested at positions -18 and -19 , as well as G to C and G to T substitutions at position -20 , immediately upstream of the T tract (Figure 4a). Finally, the entire sequence from positions -22 to -27 was substituted for an AAACCT sequence (variant mt-sp3') (Figure 4a). The effects of the mutations were tested *in vivo* using pBS::*EcogapA* plasmid derivatives. As above, the efficiency of the variant *gapA* P1 promoters was estimated by primer-extension analysis and we used the amount of *yeaA* P1 transcript for normalization of the results. Examples of cDNA fractionations are presented in Figure 4(b1). The histogram in Figure 4(b2) compares the mean values for the ratios of *gapA* P1/*yeaA* P1 transcripts that were established for the WT and variant genes in three independent experiments. For the variants showing a marked variation of *gapA* P1 activity, run-off experiments were also performed. To this end, a series of pRLG::*gapA*P1 variant plasmids was built. An example of a run-off experiment is shown in Figure 4(c1). In Figure 4(c2), we compared the mean values for the normalized amounts of *gapA* P1-*rrnB* transcript obtained for the WT and variant promoters in three independent experiments.

Not only positions but also identities of base substitutions in the T tract were found to influence promoter *gapA* P1 activity. For instance, the T to G substitution at position -16 had a strong

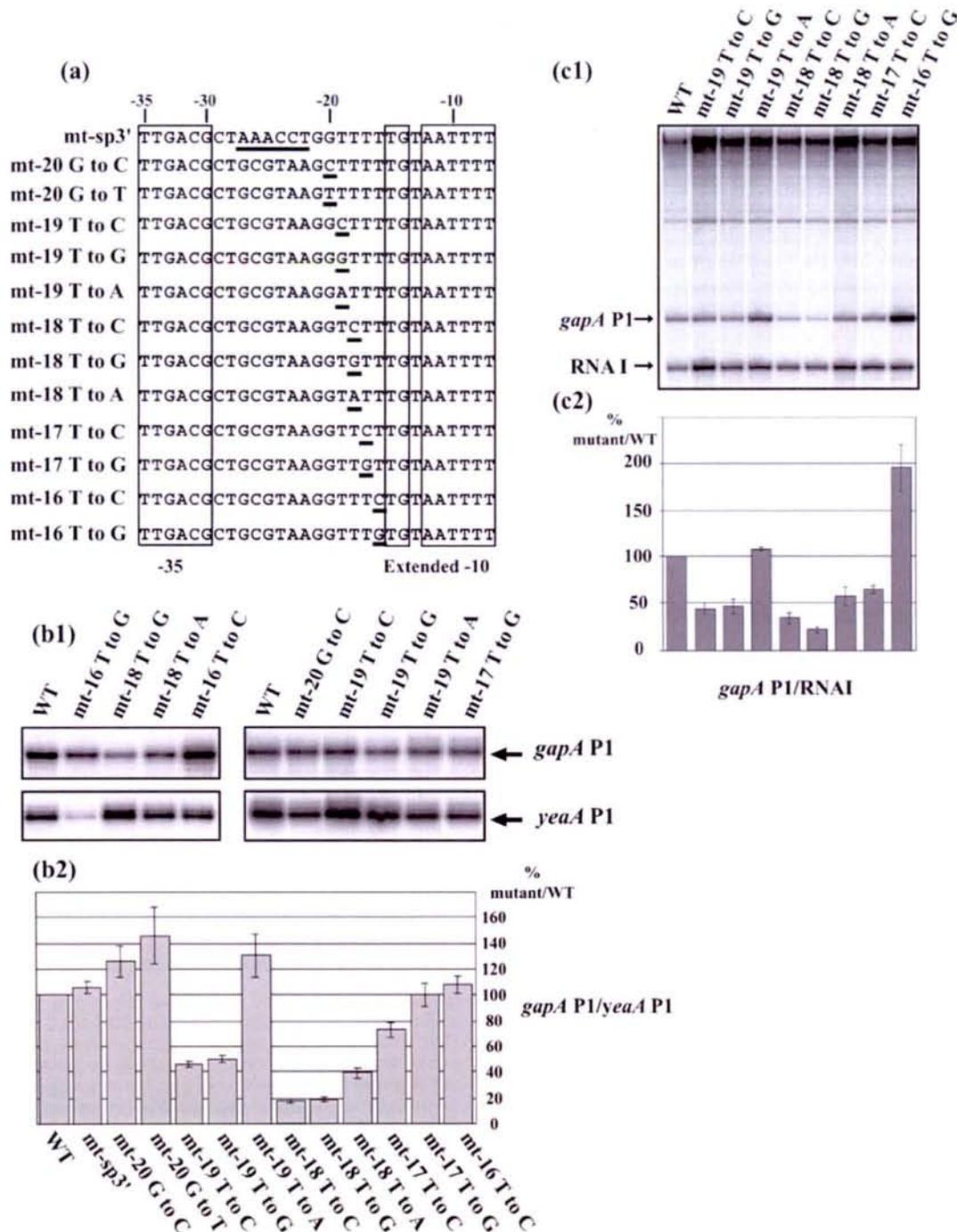


Figure 4 The sequence of the spacer region is important for *gapA* P1 activity

(a) The variant *gapA* P1 promoters with mutations in the spacer region that were produced are represented. The -35 and the extended -10 regions are boxed. The mutated residues are underlined. Names of the mutant promoters are given on the left. (b1, b2) Activities of the variant promoters were measured *in vivo* as described in Figure 1. Results are expressed as a percentage of the *gapA* P1/*yeaA* mRNA level in cells containing the WT gene. The given values are mean values for three independent experiments. The estimated error is indicated by a thin vertical bar. (c1) Plasmid pRLG::*gapA*P1 carrying the WT promoter and its derivatives, which were carrying *gapA* P1 promoters mutated in the spacer region, were used as templates for *in vitro* run-off transcription assays [28]. The transcription products were fractionated on a denaturing gel. (c2) As in Figure 3, the RNA I transcript was used for normalization of the amount of *gapA* P1 transcript recovered after each transcription experiment. Variations of the *gapA* P1 activity for the variant promoter are expressed as a percentage of the *gapA* P1/RNAI ratio found for the WT promoter (100%).

positive effect *in vitro* (factor of 2), whereas the T to C substitution at the same position only had a modest effect. Interestingly, for variant -16T to G, in agreement with the increase in *gapA* P1 activity observed *in vitro*, an increased value of the *gapA* P1/*yeaA* P1 mRNA ratio was found *in vivo*. However, the total amount of

gapA P1 mRNA in the transformed cells was low compared with that for the WT gene. This was explained by a marked decrease in the plasmid copy number. Our interpretation is that the amount of GAPDH produced from a *gapA* gene with a T to G substitution at position -16 is so high that it is deleterious for cell growth. Only

the cells that were carrying plasmids with mutations decreasing their copy number were viable. In line with this observation, a G is positioned 5' to the TG dinucleotide in variant -16T to G and the presence of a G residue at this position was already found to increase the activity of promoters from the extended -10 class [19,20]. Thus, here again, the *gapA* P1 promoter behaved as an extended -10 promoter.

For the three other T residues in the T tract, the strongest effect of base substitution was detected at position -18 (second T residue in the tract): a T to A substitution decreased the efficiency by a factor of approx. 2.5, and T to C and T to G substitutions had more severe negative effects (factor of approx. 5.5). T to C and T to G substitutions at position -19 divided the activity by a factor of approx. 2, whereas a T to A substitution at this position increased the activity. The G to C and G to T substitutions at position -20, immediately upstream of the T tract, slightly increased the promoter activity and the complete substitution of the AACCT sequence (positions -22 to -27) had a very limited effect on the activity.

Taken together, these results suggested a functional importance for the T tract sequence located in the 3' half of the spacer region, whereas the sequence of the 5' half of the spacer region did not seem to have any functional importance.

The presence of a DNA distortion in the *gapA* P1 promoter

One possible explanation for the functional importance of the 3' half of the spacer sequence was the induction of a DNA distortion by the 5-T tract that is in phase with the 4-T tract of the -10 hexamer. As DNA distortions modify DNA electrophoretic mobility, we studied the electrophoretic mobilities of DNA fragments of equal length that contained *gapA* P1 promoters. They were amplified from the WT and variant *gapA* genes by using the same pair of primers (Figure 5a). As shown in Figure 5(b), the WT DNA had an unusually low electrophoretic mobility compared with that predicted from its length (apparent mobility of 351 bp for a length of 333 bp). In addition, fragments with higher electrophoretic mobilities contained less efficient promoters, as is the case for promoter -18T to G. In contrast, the WT and mt-19T to A promoters, which have similar efficiencies, gave amplification fragments with similar low electrophoretic mobilities. These observations reinforced the idea of a DNA distortion located in the *gapA* P1 spacer that is important for promoter activity. For a better delineation of this distortion, we used the plot.it server established by Munteanu et al. [30], which allows prediction of bending of a DNA helix. As illustrated in Figure 5(c), a bend of the WT DNA is predicted in the region that contains the two series of T residues and encompasses the extended -10 sequence. In agreement with the observed electrophoretic mobilities, promoter mt-18T to G is predicted to have only a very modest bend (Figure 5c). We concluded that bending of the region, including the 3' part of the spacer region and the extended -10 element, is probably important for promoter activity.

Role of the *cis*-acting sequences for RPo formation

The question was whether DNA bending was required for RPo formation or a step further in the transcription initiation process. To answer this question, we used the KMnO_4 probing technique for the estimation of the amount of RPo formed with the variant *gapA* P1 promoters. The single-stranded thymine T-8 on the non-template strand is highly modified by KMnO_4 in the WT promoter (Figures 2 and 6) and is present in all the variant promoters. The intensity of the band corresponding to its

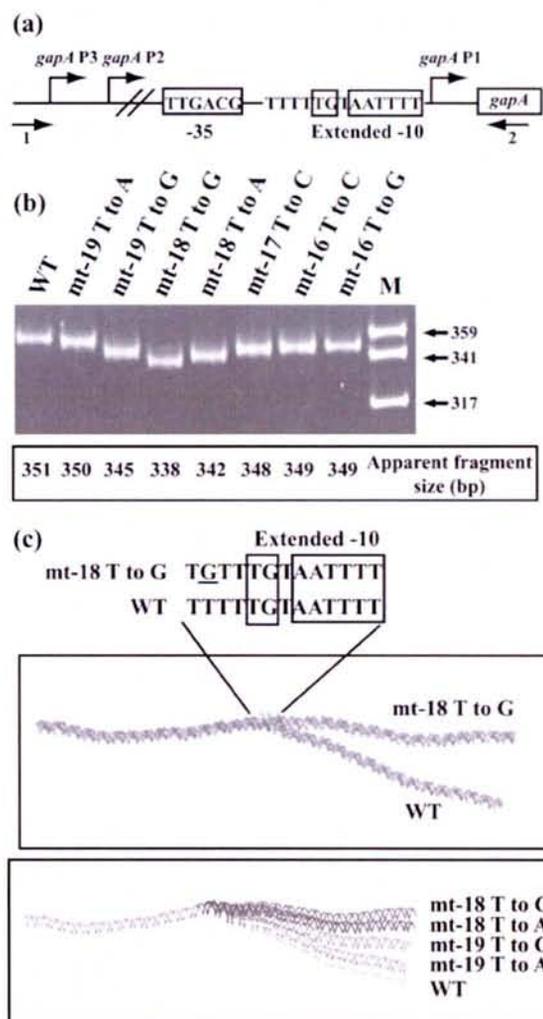


Figure 5 A DNA distortion is present in the core *gapA* P1 promoter

(a) Plasmid pBS::*EcogapA* and its derivatives were used as the templates for PCR amplifications with oligonucleotides 1 (5'-ATTGCCCTTTAAAATTCGGGG-3') and 2 (5'-TATT-CAGCACATATGCCATG-3') of fragments (+155 to -140) that carry the WT or mutated promoter sequences (see the Experimental section). (b) The amplified fragments, generated from the promoter DNA indicated above the lanes, were fractionated on a 6% polyacrylamide non-denaturing gel. The molecular-mass standard (lane M) was produced by *Sau*3AI digestion of plasmid pBR322. It was used to determine the apparent lengths of the amplified DNA fragments. (c) Models of the DNA bending of the variant *gapA* P1 promoters generated by using the plot.it server [30]. Upper panel: differential bending between WT and variant mt-18T to G *gapA* P1 promoters; lower panel: effect of a mutation at positions -18 and -19 on DNA bending.

modification by KMnO_4 was used as an indicator of the amount of RPo formed. As seen in Figure 6, no RPo formation was detected for variant mt-TG and mt-35. This demonstrated the requirement of both the TG dinucleotide and the -35 sequence for RPo formation. The amounts of RPo detected for the WT and variant mt-18T to G, mt-cons-10 and mt-cons-10/mt-35 promoters were proportional to the amounts of *in vitro* transcripts produced by these promoters (Figures 3 and 4): compared with the WT promoter, very low, similar and larger amounts of RPo were detected for the variants mt-18T to G, mt-cons-10/mt-35 and mt-cons-10 respectively. We concluded that the effects of the studied mutations took place before or at RPo formation.

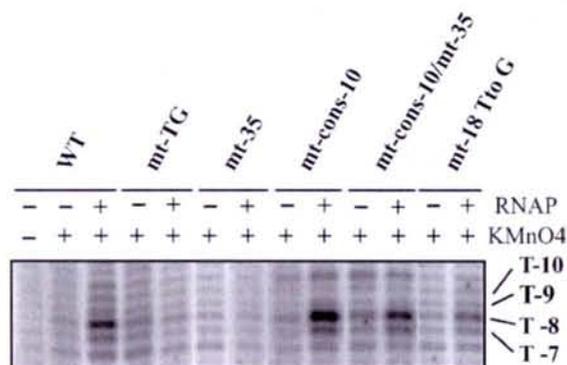


Figure 6 The TG dinucleotide, the -35 sequence and the spacer sequence of *gapA* P1 promoter are required for RPO formation

RPOs were formed with the commercial $E\sigma^{70}$ RNAP and the supercoiled plasmid pRLG::*gapA*P1 carrying the WT promoter or with the pRLG::*gapA*P1 derivatives containing the *gapA* P1 promoter variant, under conditions described in the Experimental section. RPOs were treated with $KMnO_4$ and analysed by primer extension with oligonucleotide 2559 (5'-ACCATCGGCGCTACGGCG-3') under conditions described in Figure 2. A control elongation was performed without chemical treatment (lane 1). Nucleotides with an enhanced reactivity in the RPO are indicated on the right side of the autoradiogram.

DISCUSSION

Sequence comparison of *E. coli* promoters with an extended -10 sequence suggested that some of these promoters might require a -35 sequence for activity. We bring experimental evidence for this hypothesis by an in-depth study of the *gapA* P1 promoter and performed the first chemical probing of an RPO formed at a promoter displaying this property.

$E\sigma^{70}$ RNAP selects the extended -10 non-canonical P1 instead of the canonical *gapA* P'1 promoter

The presence in the *gapA* P1 promoter region of a putative promoter (P'1) with canonical hexamers and a short spacer region (16 bp), together with a putative promoter (P1) with an extended -10 region containing a non-canonical -10 hexamer and a -35 sequence located 17 bp upstream, was an opportunity to compare the relative efficiencies of these two promoter organizations. The strong deleterious effects of the mutations at positions -14, -11 and -7 in the P1 extended -10 hexamer, as well as RPO probing data, revealed an almost unique selection of promoter P1 by $E\sigma^{70}$ RNAP (Figures 1 and 2). Both *in vivo* and *in vitro*, the putative P'1 promoter was not active or had a very low activity. Although we cannot exclude promoter P'1 utilization under different growth conditions, we concluded that the strong transcriptional activity of the *gapA* P1 region depends on a promoter that contains an extended -10 hexamer with a non-canonical -10 sequence (3/6 matches).

It was recently proposed that promoters that have a poor match to the -10 hexamer compensate by having both a TG dinucleotide and a -35 hexamer [20]. In the present study, we demonstrate that this is the case for the *E. coli gapA* P1 promoter. Indeed, the strong *gapA* P1 promoter, which plays an important role in exponentially growing cells [8], displays this kind of organization. We could, however, reinforce its activity by optimization of the -10 hexamer or by a T to G substitution at position -16. A compilation of *E. coli* promoters with putative extended -10 sequences was recently made [20]. However, owing to the occurrence of the overlapping canonical P'1 promoter, promoter *gapA* P1 was not included in the list. Thus the occurrence of a functional TG

extension among *E. coli* promoters might be more frequent than established recently.

The -12 A:T pair is opened in the *gapA* P1 RPO

Strand separation in promoter DNA induced by *E. coli* RNAP is considered to be initiated at the conserved A residue located at position -11 in the -10 hexamer. The adenine at position -11 in the non-template strand is supposed to be flipped out from the helix and tightly bound in a hydrophobic pocket of the 2.3 region of σ^{70} [33-35]; in the three-dimensional structure established for the complex formed between the *Taq* RNAP and a fork-junction DNA promoter, the T:A base-pair at position -12 is not melted [12]. A highly conserved tryptophan residue (Trp-433 in *E. coli*) of the σ 2.3 region is positioned to stack on the exposed faces of the bases at position -12. This structural arrangement is expected to delineate the upstream edge of the transcription bubble [12]. Numerous other previous experiments have strongly suggested that, for extended -10 as well as non-extended -10 promoters, the -12 T:A base-pair located at the extremity of the -10 hexamer element is not disrupted in the RPO [33,34,36,37]. In discordance with all these observations, our chemical probing of the *gapA* P1 RPO reveals an opening of the A:T base-pair, which is present at position -12, instead of the canonical T:A pair. Since only trace amounts of RPO were formed with the mt-TG variant promoter, we can exclude the possibility that a mixture of RPOs formed at the P1 and P'1 promoter sequences was present in our probing experiments (Figure 6). However, owing to the original utilization of the promoter active sequences, the base-pair at position -12 may have a lesser importance in promoter *gapA* P1 compared with other promoters studied previously. Indeed, the TA (-13 to -12) to GG (variant mt-2) substitution was found to have a limited effect, whereas the AA (-12 to -11) to GG (variant mt-9) substitution had a strong deleterious effect on *gapA* P1 activity (Figure 1). This indicates that the identity of the residue at the 5' extremity of the P1 -10 functional hexamer is not important for *gapA* P1 activity, whereas that of the residue at position -11 is essential. This functional importance of residue A-11 is in agreement with its flipping out in an RNAP pocket [33-35]. Accordingly, the second residue, which is considered to be a key recognition determinant for non-template strand binding (T at position -7) [38], is also required for efficient *gapA* P1 activity. Hence, we conclude that, as for other promoters, nucleotides -11 and -7 in the -10 hexamer are probably essential for the interaction with domain 2 of σ^{70} . However, the A:T base-pair at position -12 does not constitute the edge of the open bubble. The fact that the T-13 to G substitution (variant mt-2, Figure 1) had a very limited effect on *gapA* P1 activity excludes the possibility that the T:A base-pair at position -13 functionally replaces the T:A generally found at position -12. It should be pointed out that, in almost all RPOs previously probed by $KMnO_4$, there was a T residue at position -12 in the upper strand and this part of the strand is protected from permanganate attack by direct interaction with RNAP even if it is unwound. Thus the -12 bp opening may have escaped detection in some of the previously studied RPOs.

In *gapA* P1 RPO, the -10 element establishes numerous distinct contacts with RNAP

Under the conditions we used for RPO modification by DMS, both single-stranded A and G residues were expected to be modified, this being with a greater reactivity for G residues compared with A residues [31]. The observed pattern of modification brings information about the strong contact established between RNAP and the two single-stranded regions of the *gapA* P1 -10 hexamer,

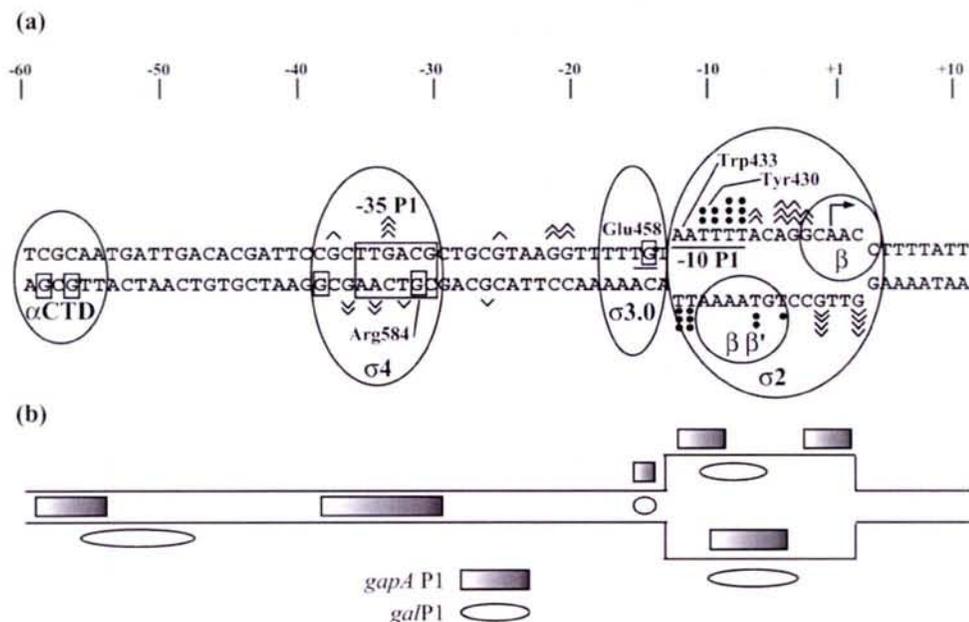


Figure 7 A complex array of DNA-protein interactions in the *gapA* P1 RPO

(a) The possible interactions of RNAP components with the upstream and core promoter regions in *gapA* P1 RPOs. Interactions of the $\sigma 2$ domain with the -10 region are in agreement with Murakami et al. [12] and Young et al. [40]. The putative interactions of residues A-11 and A-12 with Trp-433 and residues T-9 and T-10 with Tyr-430 are based on the established RF three-dimensional structure [12]. Interactions of the $\sigma 3.0$ domain with the TG extension (Glu-458) are in agreement with Barne et al. [10] and interactions between $\sigma 4$ (Arg-581) and the -35 hexamer have been described by Campbell et al. [42]. The potential interaction between α -CTD and the -60 DNA region is represented according to Ross et al. [45]. (b) Comparison of the protected areas in the *gapA* P1 RPO (the present study; rectangle) and the *galP1* RPO (ellipse) [36].

and the putative RNAP partners of the various DNA regions are indicated in Figure 7. In the non-template strand, whereas residue A-4 was strongly reactive, residue A-6 was less modified, and residues A-11 and A-12 were completely protected. Thus, in addition to residue A-11, residue A-12 is probably in close contact with factor σ^{70} (Figure 7a). The strong protection of residues A-11 and A-12 in the non-template strand is in contrast with the strong accessibility to KMnO_4 of the two opposite T residues in the template strand (Figure 2). Compared with residues T-8 and T-7, residue T-10 in the non-template strand was poorly sensitive to KMnO_4 . In the RF three-dimensional model, residues $-9/-10$ are within reach of a Tyr-283 residue (430 in *E. coli*; Figure 7). Altogether, protection of the 5' part of the -10 hexamer in the non-template strand is less extended in *gapA* P1 RPOs as compared with *galP1* RPOs (Figures 2 and 7b) [36]. It is more similar to that found for the RPO formed at the conventional *Pilv* promoter of *Bacillus subtilis* [39] and fits perfectly with crystallographic data [12]. In *gapA* P1 RPO, residues A +1 and A +2 in the non-template strand were also protected against DMS modification and their partners in the template strand were not accessible to KMnO_4 . A greater sensitivity to KMnO_4 of this region was observed in the *galP1* RPO [36] (Figure 7b) and, here again, the sensitivity found for promoter *gapA* P1 is more similar to that found for the conventional *B. subtilis Pilv* promoter [39]. According to the RF three-dimensional model, the -2 to $+4$ part of the non-template strand is held in a groove between two lobes of the β subunit [12] (Figure 7a). This may explain the observed protection. Finally, the limited accessibility of the template strand from position -5 to -10 in the *gapA* P1 RPO (Figures 2 and 7a) is in agreement with strong contacts of this DNA strand with the tunnel formed by the β' and β subunit and factor σ^{70} in the RPO [40]. Taken together, our probing data are indicative of three sites of contact between RNAP and the -10 element

(Figure 7). In addition, the observed protection of residue G-14 in the non-template strand is indicative of a direct contact of RNAP with the TG extension. Whereas residues His-455 and Glu-458 of $\sigma 3$ were found to be involved in the recognition of the -10 extension [20], no direct contact between amino acid side chains and bases at positions -14 and -15 was detected in the RF complex containing an extended -10 sequence [12]. As promoters with extended and non-extended -10 hexamers were found to form different contacts with RNAP [36,41], it will be interesting to determine which kind of interaction is responsible for protection of residue G-14 in the *gapA* P1 RPOs.

In *gapA* P1 RPO, the -35 element contacts RNAP

In the *gapA* P1 RPO, RNAP probably contacts both the -35 sequence and the TG dinucleotide, since residue G-31 in the template strand was also protected against DMS modification (Figures 2 and 7a). This protection is in perfect agreement with the interaction of residue G-31 with the Arg-409 of $\sigma 4$ (counterpart of *E. coli* Arg-584) found by X-ray analysis in the complex between *Taq* $\sigma 4$ domain and -35 DNA [42] (Figure 7a). Protection in *gapA* P1 RPO extends upstream of the -35 hexamer in the template strand (protection of residue G-38). No DNA-protein interaction was detected at this position in the complex between *Taq* $\sigma 4$ domain and -35 DNA, and the interactions with the template strand were restricted to position -31 to -33 [42]. In this complex, the non-template strand from position -35 to -38 was in contact with the $\sigma 4$ domain through protein-phosphate interactions. No protection of the corresponding segment was detected in the *gapA* P1 RPO. Hence, in *gapA* P1 RPO, RNAP may develop different contacts with the -35 region compared with those identified in the $\sigma 4$ -DNA complex [42] (Figure 7a). The increased reactivities to DMS of residues G-33 and G-37 in

the non-template strand and residues G-36, A-34 and T-32 in the template strand, which are found in *gapA* P1 RPo, are indicative of a profound distortion occurring upon RNAP binding, which is in agreement with the strong dependence of promoter *gapA* P1 activity on the -35 sequence. One important question to address is whether the interactions of $\sigma 4$ with the -35 element and $\sigma 3$ with the TG extension may occur simultaneously in the *gapA* P1 RPo. This question is important to address, since the three-dimensional model of the replication fork structure gave no information on these interactions [12]. As the function of the -35 sequence is to provide an anchor for the initial binding of RNAP [26], our results suggest that the -35 sequence plays such a role in the extended -10 *gapA* P1 promoter. The TG motif might (i) also serve as an anchor by interacting with region 3.0 of σ^{70} [10,11,13] and/or (ii) facilitate RPo formation as reported for other extended promoters [19,43] and/or (iii) stabilize the RPo as found for Gram-positive bacterial promoters [44].

The protection at positions -56 and -58 in the template strand probably represents supplementary binding sites for the RNAP (Figure 7). The bend in the -35 element, which is induced by RNAP binding, was proposed to alter the trajectory of the upstream DNA, bringing it closer to RNAP and facilitating interaction between the α -CTD (C-terminal domain of the α subunit) and upstream DNA [12]. One question that remains to be addressed is whether the α -CTD is implicated in the distal contacts detected in *gapA* P1 RPo, as found for other promoters from the non-extended class [43,45,46]. Interestingly, an upstream interaction with the RNAP holoenzyme was also detected by chemical probing of the *galP1* RPo [36] and the corresponding sequence was found to be important for promoter activity [14]. As the -35 sequence of promoter *gapA* P1 becomes dispensable when the consensus -10 sequence is present (variant mt-cons-10/mt-35 in Figure 3), the -10 hexamer, the TG motif and the upstream sequence should be sufficient for promoter recognition and stable RNAP binding.

A natural DNA distortion is important for promoter *gapA* P1 activity

Another interesting feature of promoter *gapA* P1 is the important role played by the spacer region. On the basis of the electrophoretic behaviour of a PCR-amplification product, we concluded that there is some kind of helix distortion in the core region of *gapA* P1 (Figure 5). The presence of two T tracts in phase with the B-DNA helix strongly suggests that the distortion consists of DNA bending. In addition, we found a correlation between the amplitude of the distortion and the efficiency of promoter *gapA* P1 (Figure 5), which suggests a functional importance of this natural distortion. DNA distortions were often found to be involved in the correct positioning of the -10 and -35 hexamers within RPos [26,47,48]. Residues -20 and -21, which are located immediately upstream of the T tract in the spacer region, had an increased reactivity towards DMS in the *gapA* P1 RPo. This may reflect a reinforcement of the natural DNA bending upon RNAP binding. Hence, the pre-existence of a natural DNA distortion may favour RPo formation at promoter *gapA* P1. The spacer sequence is also important for the -10 extended *galP1* promoter activity. Interestingly, the important sequence was also found to correspond to a series of T residues [14].

In conclusion, promoter *gapA* P1 is an interesting promoter that shares common properties with extended -10 promoters that do not need a -35 sequence, as well as with conventional promoters with a non-extended -10 sequence and a functional -35 sequence. For further identification of the numerous RNAP-DNA contacts we detected in the *gapA* P1 RPo, it will be interesting

to test the possibility of forming the RPo with variant RNAPs carrying mutations in the α -CTD and different regions of factor σ and of solving the three-dimensional structure of an RNAP-*gapA* P1 DNA complex. The numerous contacts established between DNA and RNAP, together with the natural capacity of bending of promoter *gapA* P1, favour a tight wrapping of the $E\sigma^{70}$ RNAP by the *gapA* P1 DNA, which probably facilitates RPo formation and explains the high efficiency of the *E. coli gapA* P1 promoter.

We are grateful to R. Gourse (University of Wisconsin) for providing us with plasmid pRLG770. M. Buckle and B. Sclavi (UMR 8113 CNRS, ENS Cachan, Cachan, France) are thanked for helpful discussions and a critical reading of this paper. G. Branlant (UMR 7567 CNRS-UHP), P. Fournieret (INIST CNRS, Nancy, France) and M. Eghbali are acknowledged for their participation in the initiation of this work. B. T. was supported by a fellowship from the French Ministère délégué à la Recherche et aux Nouvelles Technologies. This work was financially supported by the Centre National de la Recherche Scientifique (CNRS), by the French Ministère délégué à la Recherche et aux Nouvelles Technologies and by funds from the Region Lorraine (PRST Bioingénierie).

REFERENCES

- Conway, T. and Ingram, L. O. (1988) Phosphoglycerate kinase gene from *Zymomonas mobilis*: cloning, sequencing, and localization within the *gap* operon. *J. Bacteriol.* **170**, 1926-1933
- Branlant, C., Oster, T. and Branlant, G. (1989) Nucleotide sequence determination of the DNA region coding for *Bacillus stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase and of the flanking DNA regions required for its expression in *Escherichia coli*. *Gene* **75**, 145-155
- Boschi-Muller, S., Azza, S., Pollastro, D., Corbier, C. and Branlant, G. (1997) Comparative enzymatic properties of GapB-encoded erythrose-4-phosphate dehydrogenase of *Escherichia coli* and phosphorylating glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* **272**, 15106-15112
- Zhao, G., Pease, A. J., Bharani, N. and Winkler, M. E. (1995) Biochemical characterization of *gapB*-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and its possible role in pyridoxal 5'-phosphate biosynthesis. *J. Bacteriol.* **177**, 2804-2812
- Branlant, G., Flesch, G. and Branlant, C. (1983) Molecular cloning of the glyceraldehyde-3-phosphate dehydrogenase genes of *Bacillus stearothermophilus* and *Escherichia coli*, and their expression in *Escherichia coli*. *Gene* **25**, 1-7
- Branlant, G. and Branlant, C. (1985) Nucleotide sequence of the *Escherichia coli gap* gene. Different evolutionary behavior of the NAD⁺-binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* **150**, 61-66
- Charpentier, B. and Branlant, C. (1994) The *Escherichia coli gapA* gene is transcribed by the vegetative RNA polymerase holoenzyme $E\sigma^{70}$ and by the heat shock RNA polymerase $E\sigma^{32}$. *J. Bacteriol.* **176**, 830-839
- Charpentier, B., Bardey, V., Robas, N. and Branlant, C. (1998) The E11Glc protein is involved in glucose-mediated activation of *Escherichia coli gapA* and *gapB-pgk* transcription. *J. Bacteriol.* **180**, 6476-6483
- Bown, J. A., Barne, K. A., Minchin, S. D. and Busby, S. J. W. (1997) Extended -10 promoters. In *Nucleic Acids and Molecular Biology* (Eckstein, F. and Lilley, D. M. J., eds.), pp. 41-52, vol. 11, Springer, Berlin
- Barne, K. A., Bown, J. A., Busby, S. J. and Minchin, S. D. (1997) Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J.* **16**, 4034-4040
- Bown, J. A., Owens, J. T., Meares, C. F., Fujita, N., Ishihama, A., Busby, S. J. and Minchin, S. D. (1999) Organization of open complexes at *Escherichia coli* promoters. Location of promoter DNA sites close to region 2.5 of the σ^{70} subunit of RNA polymerase. *J. Biol. Chem.* **274**, 2263-2270
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. and Darst, S. A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-1290
- Sanderson, A., Mitchell, J., Minchin, S. and Busby, S. (2003) Substitutions in the *Escherichia coli* RNA polymerase σ^{70} factor that affect recognition of extended -10 elements at promoters. *FEBS Lett.* **544**, 199-205
- Chan, B. and Busby, S. (1989) Recognition of nucleotide sequences at the *Escherichia coli* galactose operon P1 promoter by RNA polymerase. *Gene* **84**, 227-236
- Keilty, S. and Rosenberg, M. (1987) Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J. Biol. Chem.* **262**, 6389-6395
- Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A. and Hayward, R. S. (1993) The minus 35-recognition region of *Escherichia coli* σ^{70} is inessential for initiation of transcription at an 'extended minus 10' promoter. *J. Mol. Biol.* **232**, 406-418

- 17 Ponnambalam, S., Webster, C., Bingham, A. and Busby, S. (1986) Transcription initiation at the *Escherichia coli* galactose operon promoters in the absence of the normal -35 region sequences. *J. Biol. Chem.* **261**, 16043-16048
- 18 Belyaeva, T., Griffiths, L., Minchin, S., Cole, J. and Busby, S. (1993) The *Escherichia coli* *cysG* promoter belongs to the 'extended-10' class of bacterial promoters. *Biochem. J.* **296**, 851-857
- 19 Burr, T., Mitchell, J., Kolb, A., Minchin, S. and Busby, S. (2000) DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Res.* **28**, 1864-1870
- 20 Mitchell, J. E., Zheng, D., Busby, S. J. and Minchin, S. D. (2003) Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res.* **31**, 4689-4695
- 21 Hawley, D. K. and McClure, W. R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* **11**, 2237-2255
- 22 Lissner, S. and Margalit, H. (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res.* **21**, 1507-1516
- 23 Auble, D. T. and deHaseth, P. L. (1988) Promoter recognition by *Escherichia coli* RNA polymerase. Influence of DNA structure in the spacer separating the -10 and -35 regions. *J. Mol. Biol.* **202**, 471-482
- 24 Spassky, A., Kirkegaard, K. and Buc, H. (1985) Changes in the DNA structure of the *lac* UV5 promoter during formation of an open complex with *Escherichia coli* RNA polymerase. *Biochemistry* **24**, 2723-2731
- 25 Kuznedelov, K., Minakhin, L., Niedziela-Majka, A., Dove, S. L., Rogulja, D., Nickels, B. E., Hochschild, A., Heyduk, T. and Severinov, K. (2002) A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition. *Science* **295**, 855-857
- 26 Record, M. T., Reznikoff, W. S., Craig, M. L., McQuade, K. L. and Schlax, P. J. (1996) *Escherichia coli* RNA polymerase (σ^{70}), promoters, and the kinetics of the steps of transcription initiation. In *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd edn (Neidhart, F., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B. and Magasanik, B., eds.), pp. 792-820, American Society for Microbiology Press, Washington, DC
- 27 Murakami, K. S. and Darst, S. A. (2003) Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13**, 31-39
- 28 Ross, W., Thompson, J. F., Newlands, J. T. and Gourse, R. L. (1990) *E. coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J.* **9**, 3733-3742
- 29 Sargent, M. G. (1968) Rapid fixed-time assay for penicillinase. *J. Bacteriol.* **95**, 1493-1494
- 30 Munteanu, M. G., Vlahovicek, K., Parthasarathy, S., Simon, I. and Pongor, S. (1998) Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.* **23**, 341-347
- 31 Borowiec, J. A., Zhang, L., Sasse-Dwight, S. and Gralla, J. D. (1987) DNA supercoiling promotes formation of a bent repression loop in *lac* DNA. *J. Mol. Biol.* **196**, 101-111
- 32 Koo, H. S., Wu, H. M. and Crothers, D. M. (1986) DNA bending at adenine-thymine tracts. *Nature (London)* **320**, 501-506
- 33 Helmann, J. D. and deHaseth, P. L. (1999) Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry* **38**, 5959-5967
- 34 Fenton, M. S. and Gralla, J. D. (2001) Function of the bacterial TATAAT -10 element as single-stranded DNA during RNA polymerase isomerization. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9020-9025
- 35 Tsujikawa, L., Strainic, M. G., Watrob, H., Barkley, M. D. and DeHaseth, P. L. (2002) RNA polymerase alters the mobility of an A-residue crucial to polymerase-induced melting of promoter DNA. *Biochemistry* **41**, 15334-15341
- 36 Chan, B., Spassky, A. and Busby, S. (1990) The organization of open complexes between *Escherichia coli* RNA polymerase and DNA fragments carrying promoters either with or without consensus -35 region sequences. *Biochem. J.* **270**, 141-148
- 37 Burns, H. D., Belyaeva, T. A., Busby, S. J. and Minchin, S. D. (1996) Temperature-dependence of open-complex formation at two *Escherichia coli* promoters with extended -10 sequences. *Biochem. J.* **317**, 305-311
- 38 Young, B. A. and Roberts, J. W. (1997) Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* **276**, 1258-1260
- 39 Juang, Y. L. and Helmann, J. D. (1994) A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids. *J. Mol. Biol.* **235**, 1470-1488
- 40 Young, B. A., Gruber, T. M. and Gross, C. A. (2002) Views of transcription initiation. *Cell (Cambridge, Mass.)* **109**, 417-420
- 41 Minchin, S. and Busby, S. (1993) Location of close contacts between *Escherichia coli* RNA polymerase and guanine residues at promoters either with or without consensus -35 region sequences. *Biochem. J.* **289**, 771-775
- 42 Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. and Darst, S. A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell* **9**, 527-539
- 43 Burns, H. D., Ishihama, A. and Minchin, S. D. (1999) Open complex formation during transcription initiation at the *Escherichia coli* *galP1* promoter: the role of the RNA polymerase alpha subunit at promoters lacking an UP-element. *Nucleic Acids Res.* **27**, 2051-2056
- 44 Voskuil, M. I. and Chambliss, G. H. (2002) The TRTGn motif stabilizes the transcription initiation open complex. *J. Mol. Biol.* **322**, 521-532
- 45 Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R. L. (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**, 1407-1413
- 46 Minakhin, L. and Severinov, K. (2003) On the role of the *Escherichia coli* RNA polymerase σ^{70} region 4.2 and a subunit C-terminal domain in promoter complex formation on the extended -10 *galP1* promoter. *J. Biol. Chem.* **278**, 29710-29718
- 47 Lozinski, T., Adrych-Rozek, K., Markiewicz, W. T. and Wierzbowski, K. (1991) Effect of DNA bending in various regions of a consensus-like *Escherichia coli* promoter on its strength *in vivo* and structure of the open complex *in vitro*. *Nucleic Acids Res.* **19**, 2947-2953
- 48 deHaseth, P. L. and Helmann, J. D. (1995) Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol. Microbiol.* **16**, 817-824

Received 12 May 2004/12 July 2004; accepted 14 July 2004

Published as BJ Immediate Publication 14 July 2004, DOI 10.1042/BJ20040792

III.) Travaux complémentaires réalisés sur le complexe RPo formé avec le promoteur *gapA* P1.

Notre étude montrait que le promoteur *gapA* P1 est tout particulièrement intéressant, puisqu'il cumule des propriétés des promoteurs classiques d'*E. coli* et des propriétés des promoteurs étendus simples qui ne nécessitent pas de boîte -35 pour être actifs. Par ailleurs, les données récentes de la littérature montrent que ce type de promoteur hybride est sans doute beaucoup plus fréquent chez *E. coli* qu'on ne l'imaginait (Mitchell *et al.*, 2003). Un effort d'étude approfondie du promoteur *gapA* P1 se justifiait donc du fait de son efficacité et des connaissances approfondies déjà acquises sur son architecture. Les expériences que nous avons réalisées *in vitro* avec l'holoenzyme $E\sigma^{70}$ purifiée montraient que l'étape limitante de l'initiation de la transcription à partir des promoteurs *gapA* P1 variants produits, dont l'efficacité était affaiblie par des mutations, est soit la formation du complexe ouvert, soit sa stabilité. En effet, pour toutes les mutations, aussi bien dans la boîte -10 étendue que dans la boîte -35 ou dans la séquence de la région interboîte, pour lesquelles nous avons constaté une diminution du taux de transcription *in vitro* et *in vivo*, nous avons constaté une diminution très marquée du taux de complexe RPo. C'est donc la formation d'un complexe RPo stable qui semble nécessiter la mise en place d'interactions multiples entre l'ADN du promoteur *gapA* P1 et l'holoenzyme $E\sigma^{70}$.

III.1) Présentation de la technique de pontage covalent par irradiation laser.

Dans l'équipe d'H. Buc, M. Buckle avait développé une méthode d'étude de la structure et de la cinétique de formation des complexes entre l'ARN pol et l'ADN, qui est basée sur la formation de pontages covalents au sein de ces complexes par irradiation avec un rayonnement laser de forte intensité, à une longueur d'onde de 266 nm (UV) (Buckle *et al.*, 1991b; Buckle *et al.*, 1999; Pemberton *et al.*, 2000; Auner *et al.*, 2003; Pemberton *et al.*, 2002). L'irradiation d'une molécule d'ADN avec ce type de laser entraîne la formation de dimères covalents de thymines proches dans l'espace ou de pontages covalents thymine-cytosine (pontages intramoléculaires). Lorsque le même type d'irradiation est réalisée sur un complexe ADN-protéine, en plus des pontages intramoléculaires, des pontages covalents sont formés entre certains acides aminés de la protéine et les bases de l'ADN (pontages intermoléculaires), ceci au niveau des zones d'interaction car seuls des résidus très proches

dans l'espace peuvent être pontés (Hockensmith *et al.*, 1986; Buckle *et al.*, 1991b; Buckle *et al.*, 1999; Pemberton *et al.*, 2000). Le taux de pontage intermoléculaire est cependant plus faible que celui de pontage intramoléculaire. Par ailleurs, la fixation de protéines sur l'ADN peut changer la conformation de la double hélice et donc modifier les pontages intramoléculaires au sein de l'ADN. En effet, la formation de dimères de thymines est très sensible à de faibles variations de conformation de la double hélice d'ADN. La comparaison des modifications générées dans l'ADN seul et dans un complexe ADN-protéine peut donc apporter des informations sur les régions d'interaction ADN-protéine, mais aussi, sur les changements de conformation de l'ADN générés par la fixation des protéines. L'avantage de cette technique est que la formation des ponts covalents est très rapide. Le temps d'irradiation est de 5 ns et le temps de la réaction photochimique est inférieur à 1 μ s (Buckle *et al.*, 1991b; Buckle *et al.*, 1999). Il est donc possible en réalisant une cinétique d'irradiation laser, de suivre l'évolution de complexes au cours du temps. La position des pontages covalents dans la molécule d'ADN est déterminée par la technique d'élongation d'amorce par le fragment de Klenow de l'ADN polymérase d'*E. coli*, que nous avons déjà cité plus haut. La formation de dimères de thymines et de pontages covalents entre l'ARN pol et l'ADN bloquent la progression de l'enzyme.

Cette technique avait déjà été utilisée par M. Buckle pour mettre en évidence les interactions ADN- $E\sigma^{70}$ au sein des complexes RPo formés avec le promoteur *lacUV5* qui est un promoteur présentant une architecture classique (Buckle *et al.*, 1991b) et avec un promoteur fort d'*E. coli* ayant une boîte -10 étendue et une boîte -35, *PtyrT* (Pemberton *et al.*, 2000; Auner *et al.*, 2003; Pemberton *et al.*, 2002). Cette technique a aussi été utilisée pour étudier la cinétique de formation des complexes promoteur-ARN pol au cours de l'initiation de la transcription à ce promoteur *lacUV5* (Buckle *et al.*, 1999). Il paraissait donc très intéressant d'essayer de l'utiliser pour compléter l'étude du promoteur *gapA* P1. Pour cela, j'ai effectué un stage dans le Laboratoire d'Enzymologie et Cinétique Structurale (LBPA, UMR 8113, Cachan) dirigé par M. Buckle.

III.2) Application de la technique de pontage covalent par irradiation laser à l'étude du promoteur *gapA* P1 sauvage.

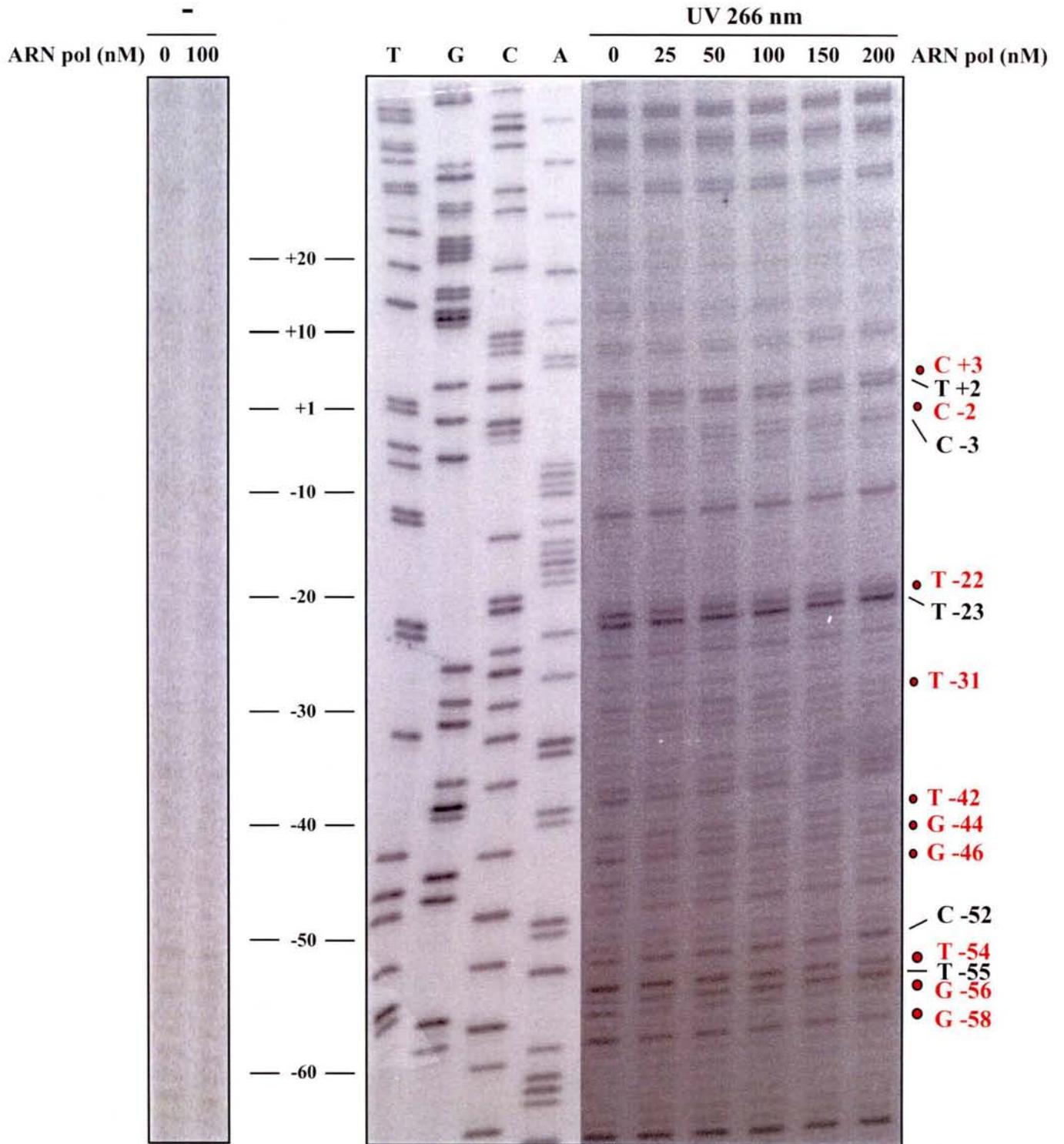
Dans un premier temps, comme cela avait été fait pour le promoteur *lacUV5*, nous avons utilisé la technique laser pour étudier le complexe RPo formé entre le promoteur *gapA* P1 et l'holoenzyme $E\sigma^{70}$ d'*E. coli*. Pour cela, nous avons formé des complexes RPo avec le

promoteur *gapA* P1 contenu dans le plasmide pRLG::*gapAPI* sous forme surenroulée. Les complexes RPo ont été formés en mélangeant le plasmide (5 nM) et l'holoenzyme, ceci avec une gamme croissante de concentrations d'enzyme allant de 0 à 200 nM. Cette enzyme avait été préparée par l'équipe de M. Buckle. L'incubation était réalisée pendant 15 min à 37°C en absence de nucléotides dans le tampon décrit dans l'article n°1. Le mélange réactionnel était ensuite soumis à un rayonnement laser (266 nm), pendant 5 ns directement dans le tube. Des amorces radiomarquées complémentaires à chacun des deux brins d'ADN du promoteur *gapA* P1 étaient alors ajoutées directement dans le mélange irradié et utilisées pour identifier par extension d'amorce avec le fragment de Klenow, la position des pontages covalents induits par l'irradiation laser. L'identification des résidus pontés dans le brin matrice et le brin non matrice sont présentés respectivement dans les figures 50 A et 50 B. L'ensemble des résultats est représenté schématiquement sur la séquence de la région promoteur dans la figure 50 C.

En absence d'ARN pol, un grand nombre de pauses sont observées le long des deux brins d'ADN. Ces pauses correspondent vraisemblablement dans beaucoup de cas à des arrêts d'élongation au niveau de paires de thymines ou de paires C-T. En particulier, sur le brin non matrice, on note une forte réactivité des deux enchaînements de T en phase avec le pas de l'hélice, qui sont présents, l'un en amont de la boîte -10 (T-16, T-17 et T-18), l'autre dans la boîte -10 (T-10, T-9 et T-8) et qui sont supposés générer une courbure de l'ADN (Fig. 50 B). La plus grande réactivité de ces résidus conforte l'idée de l'existence d'une distorsion de l'ADN. Sur le brin matrice, les résidus réactifs sont répartis tout le long de la séquence. On note néanmoins une réactivité supérieure des résidus aux positions -22 et -23 et dans la région située entre les positions -50 et -60 (Fig. 50 A).

Lors de l'ajout de quantités croissantes d'enzyme (concentrations allant de 0 à 200 nM), on s'attend dans les conditions d'incubation utilisées à avoir formation des complexes RPc, puis du complexe RPo dans lesquels la formation de dimères de thymines (ou thymine-cytosine) va être soit favorisée par changement de conformation de l'ADN permettant un rapprochement des bases, soit au contraire défavorisée par un changement de conformation entraînant leur éloignement. Par ailleurs, les contacts ADN-protéine peuvent aussi renforcer ou diminuer la réactivité des bases. Enfin, la formation de ponts covalents entre un acide aminé et l'ADN pourrait aussi être à l'origine de pauses observées lors de l'extension d'amorce, en particulier, celles détectées au niveau de résidus autres que des pyrimidines.

A



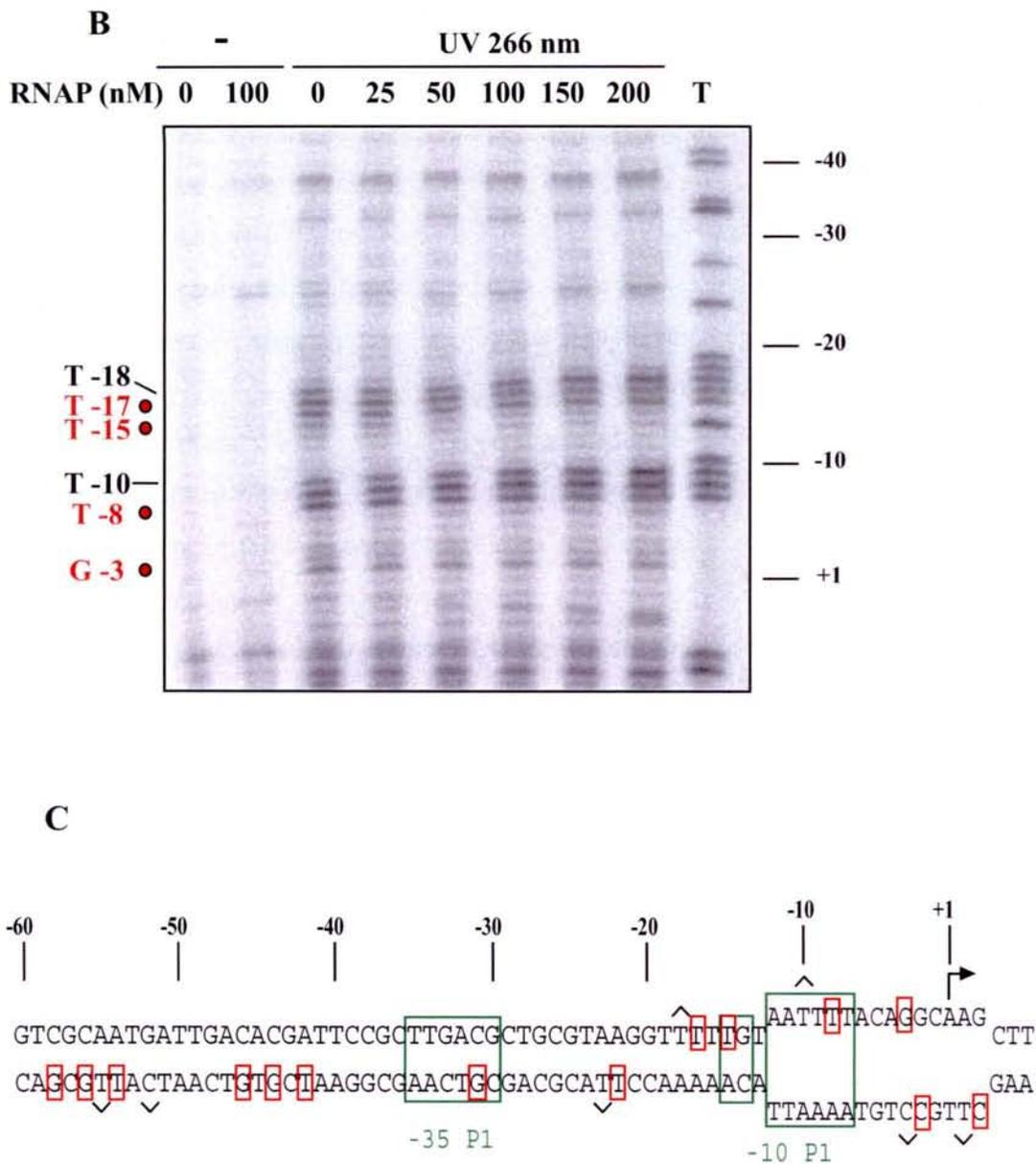


Figure 50 : Identification des résidus du promoteur *gapA* P1 ayant été le siège de la formation de ponts covalents après irradiation aux UV au sein de l'ADN nu ou de l'ADN en interaction avec l'ARN pol. L'identification a été réalisée par élévation de l'oligonucléotide 1288 par le fragment de Klenow de l'ADN polymérase d'*E. coli* sur le plasmide pRLG::*gapA*P1 irradié par un flash laser en présence de concentrations croissantes d'ARN pol. L'augmentation de l'intensité d'un arrêt d'élévation en fonction de la quantité d'ARN pol est indiquée par un trait noir. A l'inverse, une diminution de l'intensité d'un arrêt d'élévation est indiquée par un point rouge. Un séquençage réalisé avec le même oligonucléotide a été déposé à gauche. L'élévation par l'ADN polymérase s'arrête une base avant le résidu modifié. Les positions indiquées le long des gels correspondent aux positions où un pontage a eu lieu. Les complexes RPo ont été formés par incubation du plasmide et de l'ARN pol pendant 15 min à 37°C, ceci en absence de nucléotides.

A) Le brin d'ADN matrice a été analysé en utilisant l'amorce oligonucléotidique (1288) complémentaire au brin matrice.

B) Le brin non matrice a été analysé en utilisant l'amorce oligonucléotidique (2259) complémentaire au brin non matrice.

C) Représentation schématique des résultats obtenus par élévation d'amorce par l'ADN polymérase pour les deux brins d'ADN du promoteur *gapA* P1. Les boîtes -10, -35 et le dinucléotide TG sont encadrés en vert. La région de l'ADN ouverte est définie d'après les données de réactivités au DMS et au KMnO₄ décrites dans la publication n°1 (Fig. 2). Les augmentations de réactivité en fonction de la concentration en ARN pol sont indiquées par des pointes de flèche et les diminutions de réactivité par des cadres rouges.

Les variations de réactivités observées par cette approche sont en accord avec les données décrites dans l'article n°1. En effet, elles se situent au niveau de la boîte -10 étendue, de la région interboîte, de la boîte -35 et de la région amont jusqu'à la position -60 (Fig. 50 C). C'est avec le brin matrice que nous avons obtenu le plus d'informations. En aval de la boîte -10, on constate des changements du taux des pauses correspondant à 4 positions : +3, +2, -2 et -3, avec une augmentation de celles aux positions +2 et -3 et une diminution de celles aux positions +3 et -2. Cette région du promoteur est ouverte dans le complexe RPo d'après les données de l'article n°1. Dans la région interboîte, on observe une variation du taux relatif des pauses correspondant aux positions -22 et -23, avec une diminution marquée à -22 et une légère augmentation à -23, ce qui est en accord avec un changement important de la conformation de la région interboîte lors de la formation du complexe RPo. Bien que le taux des pauses observées dans l'ADN seul au niveau de la région -35 soit faible, on note une diminution de taux de certaines de ces pauses, en présence de l'ARN pol, en particulier, celle à la position -31 ce qui est en accord avec une interaction de l'ARN pol avec la boîte -35. La position -31 était aussi protégée par l'ARN pol dans les empreintes réalisées avec du DMS. Par ailleurs, de légères diminutions du taux des pauses sont aussi observées pour toute la région amont située entre les positions -42 et -46. Enfin, des variations importantes de taux sont détectées dans la région distale, en particulier, une augmentation forte de la pause à la position -55 et au contraire une diminution forte de celle à la position -54 (Fig. 50 C). Sur le brin opposé, on note surtout une augmentation de certaines des pauses dans les deux séries de T (T -10 et T -18) et la diminution d'autres pauses en présence de l'ARN pol (T-8, T-15 et T-17). Une diminution à la position G-3 est aussi observée (Fig. 50 C).

Etant donné que les modifications de réactivité observées peuvent avoir diverses origines, il est difficile d'interpréter ces données en terme de structure du complexe RPo. Un point clair qui ressort néanmoins de ces données est l'existence de changements de réactivité pour l'ensemble de la région s'étendant des positions +3 à -60, ce qui confirme l'existence de contacts nombreux entre l'ADN du promoteur *gapA* P1 et $E\sigma^{70}$ sur toute cette région d'ADN. Les modifications de réactivité des résidus T des deux enchaînements de résidus T générant la courbure naturelle pourraient refléter une augmentation ou du moins une modification du degré de cette courbure en présence de l'ARN pol. Le fort changement de réactivité des deux résidus localisés en aval de l'enchaînement de A dans le brin matrice au sein de la région interboîte pourrait aussi refléter une torsion imposée à la région interboîte par l'ARN pol. Ces résultats sont à rapprocher du modèle de structure de l'holoenzyme en association avec un

fragment d'ADN qui montre que l'ADN est courbé par l'ARN pol au niveau de la région interboîte (cf. Introduction). La même interprétation pourrait être valable pour les fortes variations détectées dans la région distale entre les positions -50 et -60 qui seraient contactées et courbées par la sous-unité α -CTD de l'ARN pol. Au contraire, les diminutions de réactivités observées pour de nombreux résidus dans la région -30 à -46 pourraient être dues à un effet écran que l'ARN pol pourrait jouer vis-à-vis de l'action du rayonnement UV sur l'ADN.

Il est intéressant de comparer nos données à celles obtenues par la même approche sur les promoteurs *lacUV5* et *PtyrT* (Buckle *et al.*, 1991b; Pemberton *et al.*, 2000; Pemberton *et al.*, 2002; Auner *et al.*, 2003), ceci d'autant plus que le promoteur *PtyrT* comporte un dinucléotide TG une paire de base en amont de la boîte -10 et pourrait donc faire partie de la famille des promoteurs étendus. Pour les promoteurs *lacUV5* et *PtyrT*, de fortes variations du taux de formation des dimères de thymines ont été observés au niveau de la boîte -35. Les variations que nous observons à ce niveau sont de faible amplitude mais ceci est vraisemblablement dû au fait que la réactivité de cette région de l'ADN *gapA* P1 est faible dans l'ADN nu.

De manière intéressante pour le promoteur *PtyrT* qui comme le promoteur *gapA* P1 renferme un dinucléotide TG, un fort changement de réactivité est observé en amont de ce dinucléotide (précisément 1 pb en amont). Nous observons le même type de changements de réactivité pour le promoteur *gapA* P1 2 pb en amont du dinucléotide TG. Ce type de réactivité n'est pas observé dans le cas du promoteur *lacUV5* et pourrait donc correspondre à un changement de conformation de l'ADN lorsque l'ARN pol $E\sigma^{70}$ se lie à un promoteur étendu.

III.3) Application de la technique de pontage covalent par irradiation laser à l'étude des variants du promoteur *gapA* P1.

La technique de pontage laser ayant donné des informations intéressantes sur le promoteur sauvage, il nous a paru intéressant de tester si elle pouvait confirmer les données obtenues sur les mutants par emploi des autres approches décrites dans l'article, voir même si elle pouvait apporter des informations complémentaires. Pour cela, nous avons sélectionné 4 variants du promoteur *gapA* P1 ayant des efficacités différentes d'initiation de la transcription. Ils sont présentés dans la figure 51 A. Les deux variants mt-35 (muté dans la boîte -35) et mt-18TtoG (muté dans la série de résidus T de la région interboîte) sont respectivement très peu et peu efficaces en transcription, alors que les deux variants mt-cons-10 et mt-cons-10/mt-35 ayant une boîte -10 correspondant à la séquence consensus sont très efficaces. Nous avons utilisé le

A

	-35 		-7 	+1 	Activité transcriptionnelle <i>in vitro</i> (% par rapport au promoteur sauvage)
Sauvage	TTGACGCTGCGTAAGGTTTTT			GTAATTTTACAGGCAACC	100
mt-18TtoG	TTGACGCTGCGTAAGGTGTTT			GTAATTTT	10
mt-cons-10	TTGACGCTGCGTAAGGTTTTTGT			TATAAT	148
mt-cons-10/mt-35	AACTCGCTGCGTAAGGTTTTTGT			TATAAT	115
mt-35	AACTCGCTGCGTAAGGTTTTT			GTAATTTT	10
	Boîte -35			Boîte -10 étendue	

gapA P1

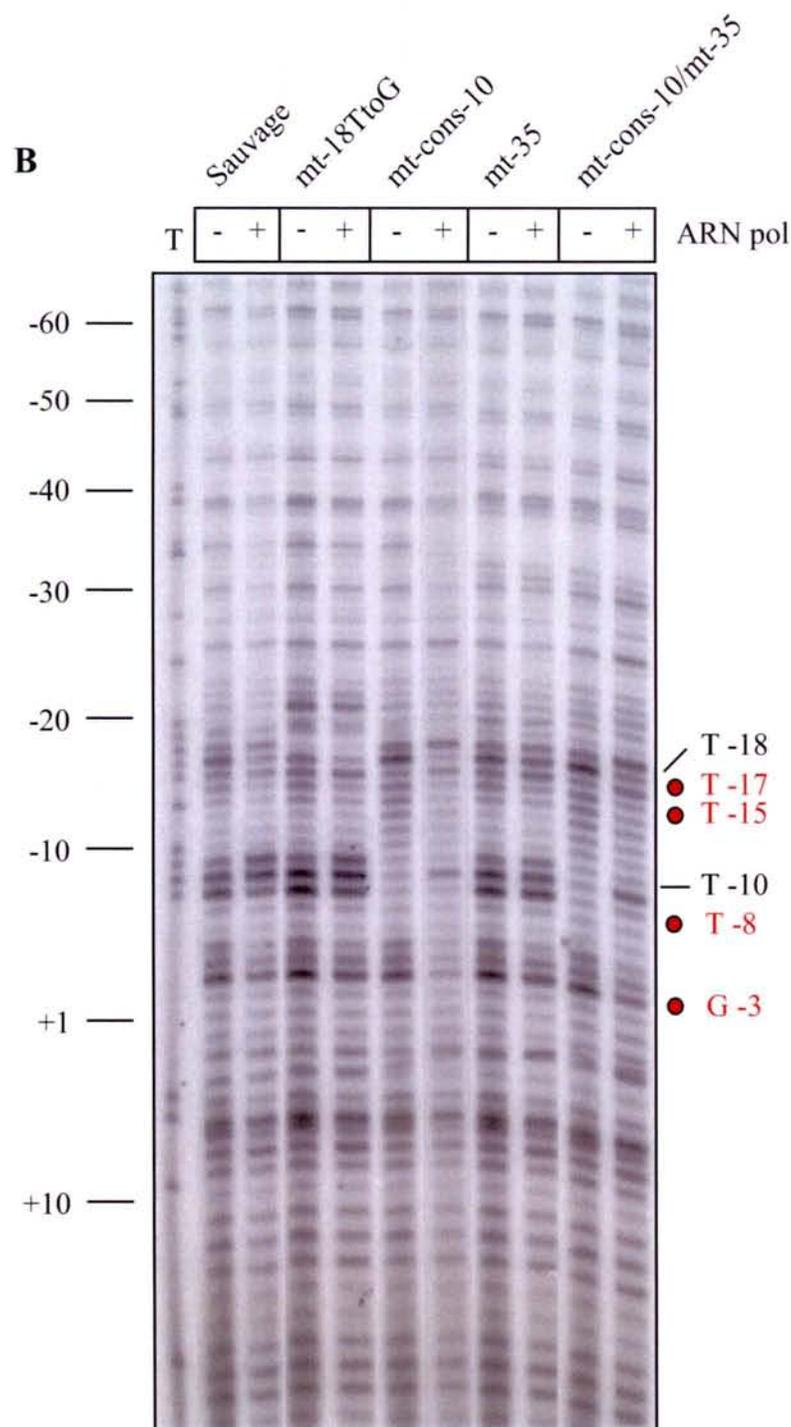


Figure 51 : Etude de la formation des complexes ouverts au niveau de promoteurs *gapA* P1 sauvage et mutés par pontage laser.

A) Séquence des variants étudiés du promoteur *gapA* P1. Les boîtes -10, -35 et le motif TG sont encadrés. Les bases qui ont été mutées sont indiquées en rouge. L'efficacité de transcription *in vitro* tirée des données de l'article n°1 est indiquée.

B) Elongation d'amorce par le fragment de Klenow de l'ADN polymérase d'*E. coli* sur le plasmide pRLG::*gapA*P1 irradié par un flash laser. Le brin non matrice a été analysé en utilisant l'amorce oligonucléotidique (2259) complémentaire au brin non matrice. Les variations de taux de formation des pontages sont représentées comme précédemment. Un séquençage a été déposé à gauche. Les positions indiquées tiennent compte du décalage de la séquence par rapport à la position de l'arrêt.

Les complexes RPo sont formés par incubation du plasmide et de l'ARN pol pendant 15 min à 37°C en absence de nucléotides.

même protocole expérimental que dans le cas du promoteur sauvage, ceci en limitant les expériences à une seule concentration d'enzyme (100 nM). Les résultats obtenus par l'analyse du brin non matrice sont présentés dans la figure 51 B. Lorsque l'on compare la réactivité des ADN seuls, les différences de réactivités correspondent essentiellement comme attendu aux variations de séquence (Fig. 52). En comparaison avec les expériences décrites ci-dessus, on note une différence de profil pour le complexe RPo formé entre l'ARN pol et le promoteur sauvage. Cette différence de réactivité s'explique sans doute par un changement de lot d'ARN pol. Un point intéressant apparaît lors de la comparaison des réactivités de l'ADN sauvage et des ADN mutés en présence de l'ARN pol. Pour les deux variants peu efficaces en transcription mt-18TtoG et mt-35, alors qu'une modification de la réactivité est observée dans la région -15 à -19, les changements de réactivité au niveau de la série de résidus T de la boîte -10 sont soit beaucoup plus faibles que pour l'ADN sauvage (mt-18TtoG), soit inexistant dans le cas du variant mt-35, ce qui suggère une fixation possible de l'ARN pol sans que cette fixation aboutisse à la formation du complexe ouvert (mt-35) ou du moins qu'elle aboutisse à un faible taux de complexe ouvert (mt-18TtoG). Ces données sont en parfait accord avec celles obtenues par l'emploi du KMnO_4 montrant que le taux de complexe RPo formé avec ces deux variants est nettement plus faible qu'avec l'ADN sauvage.

Pour le variant mt-cons-10 qui est plus efficace et pour le variant mt-cons-10/mt-35, qui est aussi efficace que le promoteur sauvage, nous observons de très fortes variations des réactivités dans la région allant des positions -18 à -3 en présence de l'ARN pol. On note en plus une augmentation de la réactivité à la position -26 (Fig. 51 B et 52).

III.4) Etude de la stabilité des complexes RPo formés au promoteur *gapA* P1.

Etant donné que l'approche laser et l'analyse des complexes RPo nous conduisaient à conclure que toutes les mutations délétères pour l'activité du promoteur *gapA* P1 conduisent soit à un ralentissement de la cinétique de formation du complexe RPo, soit à son instabilité, il nous a paru intéressant d'essayer d'utiliser l'approche laser pour apporter des informations sur ces deux explications possibles. Cette approche peut en effet, permettre d'évaluer la stabilité des complexes RPo. Pour cela, de l'héparine doit être ajoutée dans le milieu réactionnel où est formé le complexe ouvert. L'héparine se fixe dans le canal de fixation de l'ADN dans l'ARN pol libre et ainsi, l'ARN pol ne peut plus se fixer sur l'ADN promoteur et donc former des complexes RPo. L'héparine entre aussi en compétition avec l'ADN associé à l'ARN pol au sein des complexes RPo, ce qui entraîne leur dissociation. Seuls les complexes RPo restent

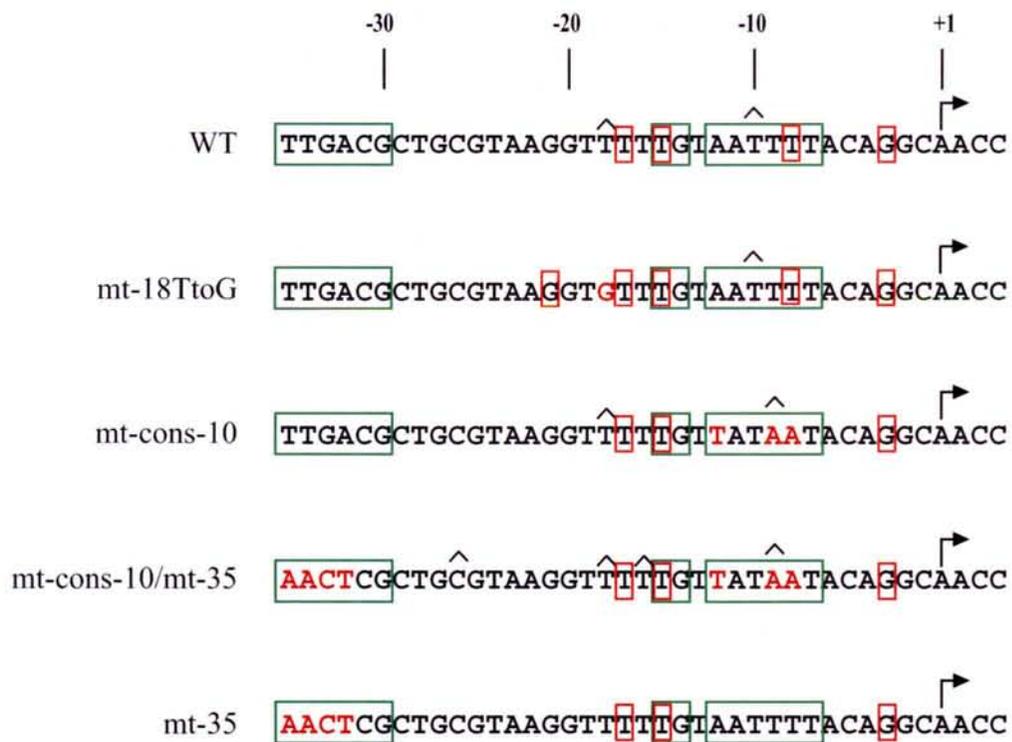
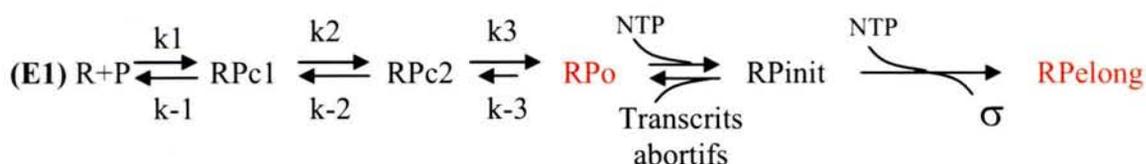


Figure 52 : Représentation schématique des résultats obtenus par élongation d'amorce par l'ADN polymérase pour le brin non matrice des variants du promoteur *gapA* P1. Les boîtes -10, -35 et le dinucléotide TG sont encadrés en vert. Les augmentations de réactivité en fonction de la concentration en ARN pol sont indiquées par des pointes de flèche et les diminutions de réactivité sont indiquées par des cadres rouges. Les bases qui ont été mutées sont indiquées en rouge.

stables aux concentrations d'héparine utilisées pour ce type d'expérience (25 µg/ml). Donc au cours d'une incubation avec l'héparine, les complexes RPo se dissocient plus ou moins vite en fonction de leur stabilité, libérant de l'ARN pol libre, qui est bloquée par l'héparine. Cette réaction de dissociation est régie par la constante de vitesse k_3 de l'équation (E1) de formation des complexes d'initiation de la transcription.



Il est possible d'avoir une estimation de cette constante si l'on dispose d'un signal généré par irradiation laser qui permette d'évaluer la quantité de complexe RPo présent dans le mélange. Les expériences décrites ci-dessus pour le promoteur sauvage et les promoteurs variants (Fig. 51 B) mettaient en évidence un changement marqué de la sensibilité du résidu T-17 au rayonnement UV lors de la formation du complexe RPo, ceci dans le promoteur sauvage et les deux promoteurs variants *mt-cons-10* et *mt-cons-10/mt-35* et à un moindre degré pour le variant *mt-18TtoG*. Nous avons donc décidé d'utiliser ce signal comme signature de la présence du complexe RPo. Pour cela, les complexes RPo ont été formés avec le promoteur *gapA P1* comme précédemment, puis après 15 min d'incubation à 37°C, nous avons ajouté de l'héparine. Le moment de l'ajout correspond au temps 0. Des échantillons du mélange réactionnel ont été prélevés après des temps d'incubation de 1, 2, 3, 5, 10, 30 et 60 min après ajout de l'héparine et ils ont été soumis immédiatement à une irradiation laser. L'ADN de ces échantillons a ensuite été analysé par élongation d'amorce par l'ADN pol comme décrit ci-dessus.

Les résultats que nous avons obtenus pour le promoteur sauvage ainsi que pour les variants *mt-cons-10*, *mt-18TtoG* et *mt-cons-10/mt-35* sont présentés dans la figure 53. D'après les résultats ci-dessus, pour l'ADN sauvage et les deux variants *mt-cons-10* et *mt-cons-10/mt-35*, après 15 min d'incubation entre l'ADN et l'holoenzyme et lorsque l'héparine est ajoutée, l'ADN est majoritairement sous forme de complexe RPo. Ceci se traduit par une forte diminution du signal à la position T -17. L'effet est beaucoup moins marqué pour le variant *mt-18TtoG* pour lequel un faible taux de complexe RPo est présent après ce temps d'incubation.

Pour l'ADN sauvage et le variant *mt-cons-10/mt-35*, le signal à la position T-17 réapparaît au cours du temps (Fig. 53), ce qui doit refléter la dissociation du complexe RPo. Les données

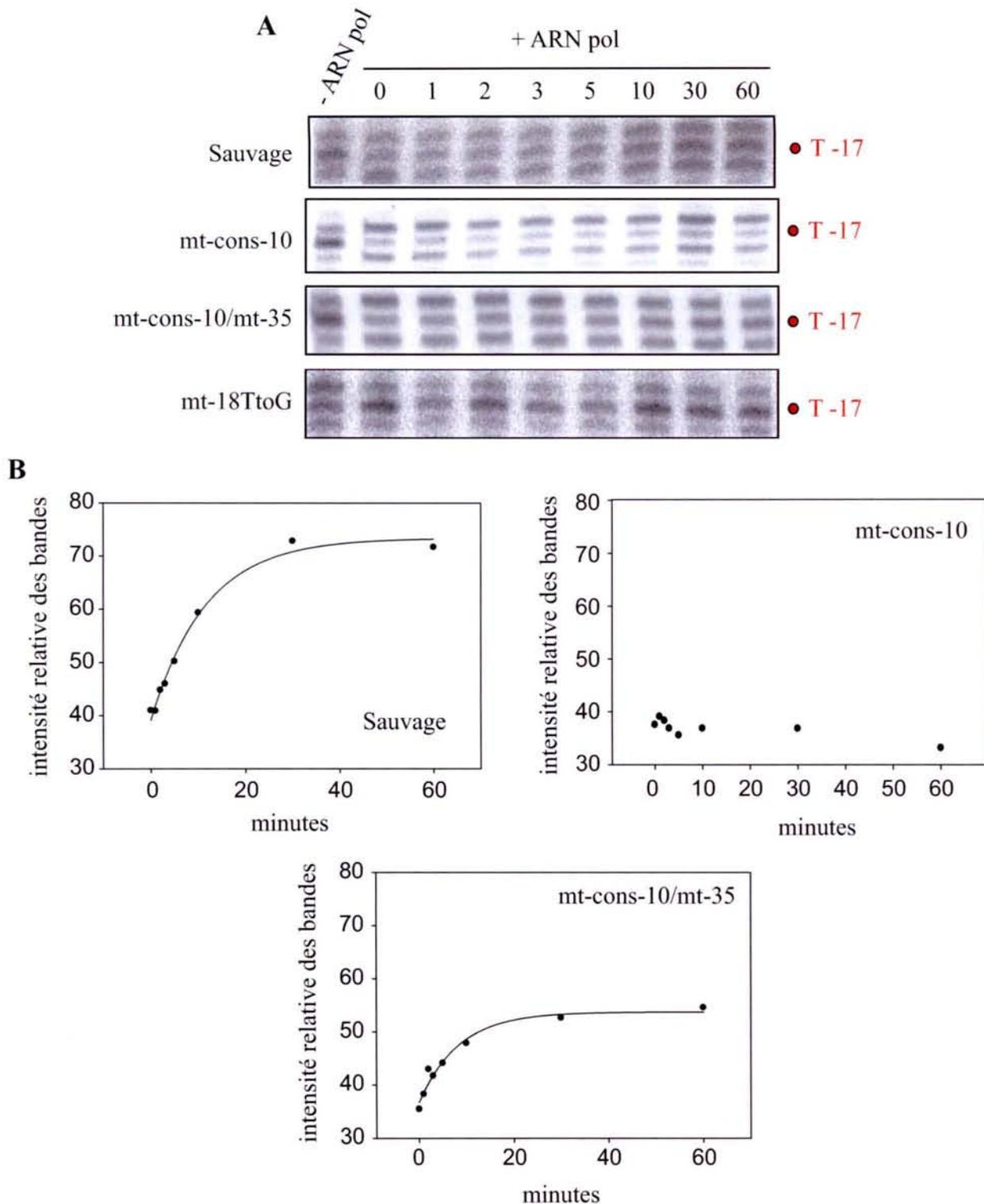


Figure 53 : Etude de la stabilité des complexes RPo formés au niveau du promoteur *gapA* P1 et des variants mt-cons-10 et mt-cons-10/mt-35 par formation de pontages covalents par laser UV.

A) Elongations d'amorce par le fragment de Klenow de l'ADN polymérase réalisées sur le brin non matrice du promoteur *gapA* P1 sauvage ou variant après irradiation aux UV. La réactivité à la position T-17 a été utilisée comme signal représentatif de la diminution du taux de complexes RPo. Le temps 0 est le moment à partir duquel l'héparine a été ajoutée (15 min après l'incubation de l'holoenzyme avec l'ADN à 37°C) et les temps indiqués au-dessus des pistes correspondent au moment où l'échantillon a été prélevé et soumis au rayonnement laser. La première piste correspond à l'élongation d'amorce en absence d'ARN pol.

B) Apparition de la réactivité de la base T-17 en fonction du temps. Les signaux en A) ont été quantifiés ainsi qu'un signal ne variant pas en présence de l'ARN pol (référence). Le rapport entre ces deux signaux a été calculé. Le pourcentage que cette valeur représente par rapport à la valeur obtenue pour l'expérience réalisée en absence d'ARN pol est portée en ordonnée et le temps en abscisse. Les données obtenues ont été exploitées en considérant que leurs variations correspondent à l'équation $y=y_0+A(1-e^{-kt})$ et en utilisant le logiciel Sigma plot (SPSS Inc.).

obtenues suggèrent que cette dissociation est très faible pour le variant mt-cons-10. En effet, le signal en position -17 ne réapparaît pas même après 60 min d'incubation, ce qui suggère une forte stabilité du complexe RPo formé avec ce variant. Nous avons quantifié pour chaque échantillon, la radioactivité présente dans la bande T-17, ainsi que celle d'une bande dont l'intensité de radioactivité ne semblait pas varier lors de la formation des complexes (position +15 hors du promoteur). Le rapport des valeurs obtenues pour chacune des bandes a été calculé pour chaque échantillon et ce rapport a ensuite été exprimé sous la forme d'un pourcentage par rapport à la valeur obtenue pour l'échantillon sans ajout d'ARN pol. La courbe représentant les variations de ces pourcentages en fonction du temps auquel l'échantillon a été prélevé, a été tracée (Fig. 53 B). Ces variations répondent à l'équation mono exponentielle $y = y_0 + A(1 - e^{-kt})$ (Pemberton *et al.*, 2002) dans laquelle, la constante de vitesse k représente la constante de vitesse de dissociation du complexe RPo. La valeur de k obtenue dans le cas du promoteur sauvage est de $0,086 \pm 0,012 \text{ min}^{-1}$ (Tableau 4).

Promoteur <i>gapA</i> P1	Activité transcriptionnelle <i>in vitro</i> (% par rapport au promoteur sauvage)	Constante de vitesse de dissociation du complexe RPo (min^{-1})
Sauvage	100	$0,086 \pm 0,012$
mt-cons-10	148	Très faible (10^{-3})
mt-cons-10/mt-35	115	$0,121 \pm 0,028$

Tableau 4 : Bilan des résultats des études cinétiques réalisées pour le promoteur *gapA* P1 sauvage et les variants mt-cons-10 et mt-cons-10/mt-35. Le calcul des constantes de vitesse est décrit dans le texte et l'activité transcriptionnelle est décrite dans la publication.

Pour le variant mt-cons-10, la valeur de k obtenue est très faible (de l'ordre de 10^{-3}) par rapport à celle du promoteur sauvage (Tableau 4). Nous pouvons ainsi conclure que lorsque la séquence de la boîte -10 correspond à la séquence consensus, le complexe RPo est très stable (environ deux log d'écart entre les constantes de vitesse de dissociation). Cette stabilité est en accord et explique sans doute l'augmentation de l'efficacité du promoteur *gapA* P1 observée *in vitro* (cf. article, Fig. 3).

In vitro, le promoteur variant mt-cons-10/mt-35 initie la transcription avec une efficacité semblable à celle du promoteur sauvage (Tableau 4). La constante de vitesse de dissociation k pour ce variant calculée à partir de la réapparition du signal T-17 est de $0,121 \pm 0,028 \text{ min}^{-1}$.

Cette valeur est du même ordre de grandeur que celle obtenue pour le promoteur sauvage. Nous concluons donc que ces deux promoteurs, avec des architectures différentes, possèdent des efficacités d'initiation de la transcription similaires, ce qui est sans doute lié au fait que la stabilité des complexes RPo formés au niveau de ces promoteurs est semblable. Il est à noter que dans le cas du variant mt-cons-10/mt-35, la courbe n'atteint pas la même valeur plateau que pour l'ADN sauvage, ce qui indique que la réactivité du résidu thymine à la position T-17 n'est pas la même avant formation du complexe RPo et après sa dissociation. Cette différence pourrait être expliquée par le fait que la dissociation du complexe RPo passe par des complexes intermédiaires non observés dans le cas du promoteur sauvage. On peut imaginer que l'ARN pol reste fixée à l'ADN, maintenant ce dernier dans une conformation ne conduisant pas à une augmentation du signal à la position -17 jusqu'à la valeur obtenue pour l'ADN de départ. Nous ne pouvons pas non plus exclure la formation possible de deux types de complexes RPo au niveau de ce promoteur variant, l'un très stable et l'autre ayant la même stabilité que celui formé avec le promoteur sauvage. Des expériences complémentaires sont nécessaires pour trancher entre les différentes explications possibles.

D'après ces résultats, nous pouvons supposer que dans le cas du promoteur *gapA* P1, plus les complexes RPo sont stables, plus l'initiation de la transcription est efficace. En effet, la mutation de la séquence de la boîte -10 conduit à une augmentation de l'efficacité transcriptionnelle *in vitro* mais aussi à la stabilisation des complexes RPo.

Les résultats obtenus pour le variant mt-18TtoG sont très difficilement interprétables en l'état, des expériences complémentaires devront être réalisées.

Les résultats présentés dans la figure 6 de la publication nous amènent à conclure que les mutations qui limitent ou améliorent l'efficacité de l'initiation de la transcription au promoteur *gapA* P1 conduisent à un taux de complexe RPo, respectivement plus faible ou plus élevé. Les données obtenues par l'approche laser suggèrent que la diminution ou l'augmentation du taux de complexes RPo dépendrait beaucoup de la stabilité de ces complexes. Pour aller plus loin dans l'analyse du mécanisme d'initiation de la transcription au promoteur *gapA* P1, il sera intéressant de définir les constantes de vitesse de formation des différents intermédiaires comme cela a été réalisé dans le cas du promoteur *lacUV5* (Buckle *et al.*, 1999). Pour cela, des signaux caractéristiques de la formation des différents complexes RPo1, RPo2 et RPo devront être identifiés et quantifiés lors de cinétiques rapides. En effet, la formation de ces complexes est très rapide. Afin d'étudier ces intermédiaires, de faible durée de vie, les expériences pourront être réalisées à basse température (22°C).

III.5) L'efficacité du promoteur *gapA* P1 dépend du taux de surenroulement de l'ADN.

Nos données montraient que l'efficacité du promoteur *gapA* P1 dépend de sa séquence mais aussi de la présence d'une courbure de l'ADN. Il était donc intéressant de tester si le surenroulement de l'ADN qui peut aussi modifier la conformation de l'ADN, influence l'initiation de la transcription au promoteur *gapA* P1 (Fig. 54). B. Charpentier avait déjà comparé les efficacités de transcription *in vitro* du promoteur *gapA* P1 présent dans un plasmide surenroulé ou linéarisé par clivage par une enzyme de restriction. Les résultats présentés dans la figure 54 A indiquent qu'*in vitro*, une initiation de la transcription efficace n'est obtenue que lorsque le promoteur *gapA* P1 est porté par un plasmide surenroulé. *In vitro*, l'efficacité du promoteur *gapA* P1 dépend donc du surenroulement de l'ADN.

Pour valider cette hypothèse par une étude *in vivo*, j'ai utilisé des souches d'*E. coli* mutées l'une dans le gène codant la topoisomérase (JTT1 *topA*⁻), l'autre dans le gène codant une sous-unité de l'ADN gyrase (JTT1 *gyrB*⁻). Le taux de surenroulement négatif est plus élevé dans la souche *topA*⁻ et plus bas dans la souche *gyrB*⁻ (Steck *et al.*, 1993). Après culture de ces souches en milieu LB et extraction des ARN totaux, j'ai analysé le taux de transcrits initiés au promoteur *gapA* P1 par élongation d'amorce à la transcriptase inverse, ceci en utilisant le protocole décrit dans la publication n°1. Les résultats montrent qu'*in vivo*, un taux de surenroulement élevé conduit à une diminution du taux de transcrits initiés au promoteur *gapA* P1, alors qu'un taux de surenroulement plus bas conduit à une augmentation du taux de transcrits initiés à ce promoteur (Fig. 54 B). Il est à noter que cet effet observé *in vivo* n'est pas forcément direct mais il va dans le même sens que les effets observés *in vitro*, ce qui suggère fortement que le promoteur *gapA* P1 est sensible au taux de surenroulement de l'ADN. Il est à noter que lors d'études de la variation du taux des protéines d'*E. coli* en fonction du degré de surenroulement de l'ADN par électrophorèse 2D, il avait été constaté que la GAPDH varie avec le taux de surenroulement, ce qui va dans le même sens que nos données *in vitro* et *in vivo* (Steck *et al.*, 1993).

En conclusion, les éléments de séquence importants pour l'initiation de la transcription au promoteur *gapA* P1, influencent l'étape de formation du complexe RPo. Les nombreuses interactions entre l'ARN pol et le promoteur semblent nécessaires pour permettre la formation de complexes ouverts stables conduisant à une initiation de la transcription efficace. La

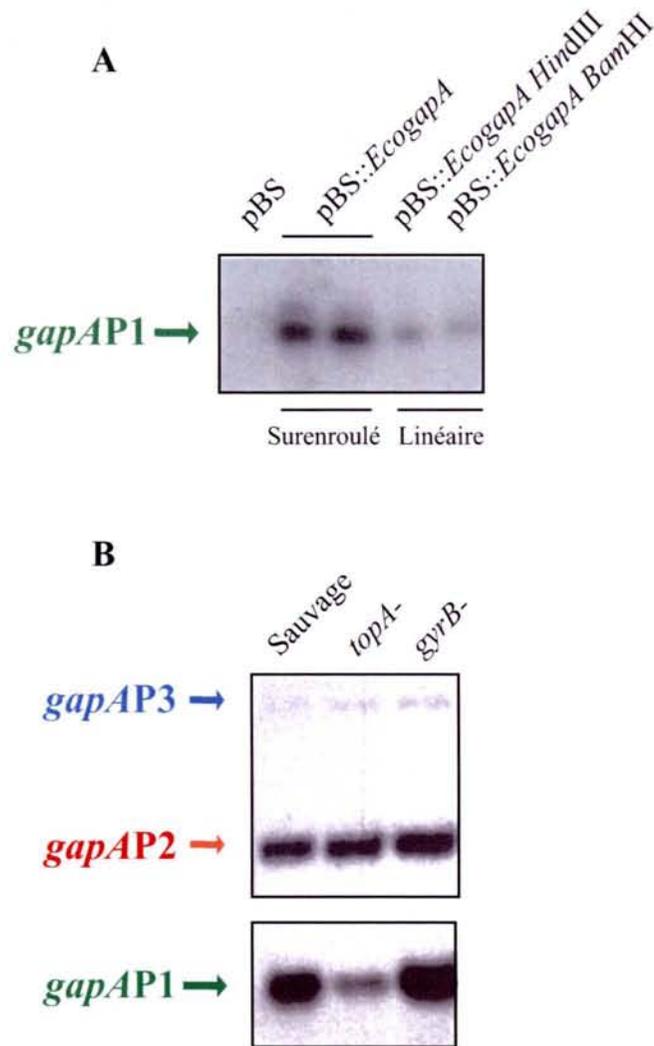


Figure 54 : Effet du taux de surenroulement de l'ADN sur l'initiation de la transcription au promoteur *gapA* P1.

A) Transcription *in vitro* réalisée avec l'holoenzyme $E\sigma^{70}$ purifiée en utilisant comme matrice le plasmide pBS::EcogapA (cf. article n°1) sous forme surenroulée ou linéaire. La matrice linéaire a été produite par coupure du plasmide par les enzymes indiquées en haut des pistes. Le plasmide pBS a été utilisé comme témoin (piste 1). Le protocole utilisé est décrit dans l'article n°1.

B) Analyse des taux de transcrits *gapA* P1, *gapA* P2 et *gapA* P3 par élongation d'amorce par la transcriptase inverse. Les souches JTT1, JTT1 *topA*⁻ et JTT1 *gyrB*⁻ ont été cultivées en milieu LB et les cellules ont été prélevées à une DO_{600nm} de 0,5. Les ARN totaux ont été extraits par la méthode au phénol chaud, puis 20 µg d'ARN totaux ont été analysés par extension d'amorce par la transcriptase inverse. L'amorce utilisée était complémentaire aux ARNm transcrits à partir des promoteurs *gapA* P1, P2 et P3. Les ADNc ainsi obtenus ont été fractionnés en gel de polyacrylamide dénaturant. La nature de l'ADNc est indiquée à gauche du gel.

présence d'une déformation interne de la séquence promoteur et le surenroulement de l'ADN doivent favoriser la mise en place de l'ensemble de ces interactions.

D'après nos résultats, la séquence de la boîte -10 du promoteur *gapA* P1 est sous-optimale et bien que le promoteur *gapA* P1 ait une activité constitutive efficace, l'initiation de la transcription au niveau de ce promoteur peut encore être améliorée en mutant la séquence de la boîte -10 pour qu'elle corresponde au consensus reconnu par $E\sigma^{70}$. Cela signifie qu'au cours de l'évolution, la séquence du promoteur n'a pas été optimisée de façon à engendrer une activité transcriptionnelle maximale. Cette propriété pourrait permettre l'activation par des facteurs de régulation permettant de moduler le taux d'utilisation du promoteur *gapA* P1 en fonction des conditions de croissance. Ces régulations pourraient permettre de relier l'efficacité du promoteur *gapA* P1 aux besoins de la cellule. Nous décrirons par la suite les différents éléments qui pourraient réguler l'initiation de la transcription au niveau de ce promoteur.

Partie 2 : Etude de l'ensemble de la région intergénique *gapA-yeaA*.

I.) Intérêt d'une étude de la région intergénique *gapA-yeaA*.

Comme nous l'avons déjà indiqué, le promoteur *gapA* P1 est l'un des 4 promoteurs pouvant diriger la transcription de l'ORF *gapA*. B. Charpentier avait détecté un certain nombre de particularités de la région contenant ces 4 promoteurs, en particulier, la présence de plusieurs sites de fixation de la protéine Fis. Les uns sont localisés en amont du promoteur *gapA* P1, les autres sont présents en amont du promoteur *gapA* P4. La comparaison du taux de transcrits initiés au promoteur *gapA* P1 dans une souche *fis*⁻ et dans une souche non isogénique *fis*⁺, qu'avait réalisée B. Charpentier, suggérait un rôle activateur de la protéine Fis, vis-à-vis du promoteur *gapA* P1, mais ces données n'avaient pas été approfondies.

B. Charpentier avait aussi mis en évidence l'existence d'une courbure de forte amplitude dans la région située en amont du promoteur *gapA* P3, courbure qui chevauche un site de fixation du complexe CRP-AMPc. Un mutant qui touchait à la fois la courbure et le site de fixation du complexe CRP-AMPc avait révélé l'importance de ce segment de la région inter-ORF sur le taux d'utilisation des promoteurs *gapA* P2 et *gapA* P3. Mais, dans ce cas aussi, les données n'avaient pas été approfondies.

Enfin, le traitement par KMnO₄ de complexes formés entre l'ADN de la région intergénique *yeaA-gapA* et l'holoenzyme Eσ⁷⁰ avait permis de mettre en évidence la formation d'un complexe ouvert en amont de l'ORF *yeaA*, révélant la présence d'un promoteur, actif *in vitro* sans activateur. A cette époque, l'identité du gène *yeaA* n'était pas connue et l'étude n'avait pas été poursuivie. La connaissance de l'identité de la protéine codée par le gène *yeaA* (la protéine MsrB) et son importance pour les mécanismes de résistance de la bactérie au stress oxydant, nous a conduit à poursuivre l'étude de la séquence initiant la transcription de ce gène. Ceci d'autant plus qu'on avait à faire à une région intergénique très intéressante puisqu'elle renferme 4 promoteurs initiant la transcription du gène *gapA* sur l'un des deux brins et deux promoteurs transcrivant *yeaA* sur le brin opposé, ceci avec la présence de nombreux sites potentiels de fixation pour la protéine Fis, d'une forte courbure de l'ADN et d'un site de fixation du complexe CRP-AMPc, tous ces éléments pouvant moduler la transcription de l'un, de l'autre ou des deux gènes *yeaA* et *gapA*.

En utilisant l'approche transcription *in vitro* décrite dans l'article n°1 et basée sur l'emploi du plasmide pRLG770 ou ses dérivés et de l'holoenzyme $E\sigma^{70}$ purifiée et par analyse du taux d'ARNm produits *in vivo* par la technique d'extension d'amorce par la transcriptase inverse, aussi décrite dans l'article, j'ai pu montrer en produisant différents variants de la région inter-ORF *yeaA-gapA*, que plusieurs des éléments particuliers portés par cette séquence inter-ORF influencent le taux de transcription des deux gènes *gapA* et *yeaA* et souvent de façon opposée, comme nous allons le décrire.

II.) Rôle des sites de fixation de la protéine Fis dans la régulation de la transcription des ORF *gapA* et *yeaA*.

L'analyse de la région inter-ORF *gapA-yeaA* met en évidence l'existence de 5 sites potentiels de fixation de la protéine Fis. La séquence consensus établie pour les sites de fixation de Fis est en effet GNtYAa(A/T)(A/T)(A/T)tTRaNC, où N représente A, C, G, T ; R : A, G ; Y : C, T. Les lettres en majuscule correspondent à un niveau de fréquence élevé des résidus correspondants et les lettres en minuscule correspondent à une fréquence plus faible des résidus (Hengen *et al.*, 1997) (Fig. 45 et 55) (Charpentier, 1994). Les deux sites potentiels de fixation de la protéine Fis situés en amont du promoteur *gapA* P1 sont, respectivement, centrés aux positions -62 (site A) et -68 (site B) par rapport au site d'initiation de la transcription. Le site A pourrait être un site de fixation d'un activateur de classe I. Il est à noter que la séquence du site A chevauche partiellement le segment situé aux positions -50 à -60 qui est protégé contre l'action du DMS par l'ARN pol dans le complexe RPo (Fig. 2 article n°1). Par ailleurs, nous avons aussi observé que la réactivité de ces résidus vis-à-vis du rayonnement UV est modifiée. Trois autres sites potentiels de fixation de la protéine Fis sont détectés plus en amont du promoteur *gapA* P4 et nous présenterons leur étude un peu plus loin dans cette partie du chapitre résultats.

II.1) Etude de l'effet de la protéine Fis sur l'activité du promoteur *gapA* P1.

Nous commencerons par présenter l'étude de l'effet de la protéine Fis sur le promoteur *gapA* P1. En utilisant la souche CSH50*fis::kan* (Koch *et al.*, 1988) et une souche non isogénique, toutes deux transformées par le plasmide pBS::*EcogapA*, B. Charpentier avait mis en évidence par extension d'amorce à la transcriptase inverse, une légère diminution du taux de

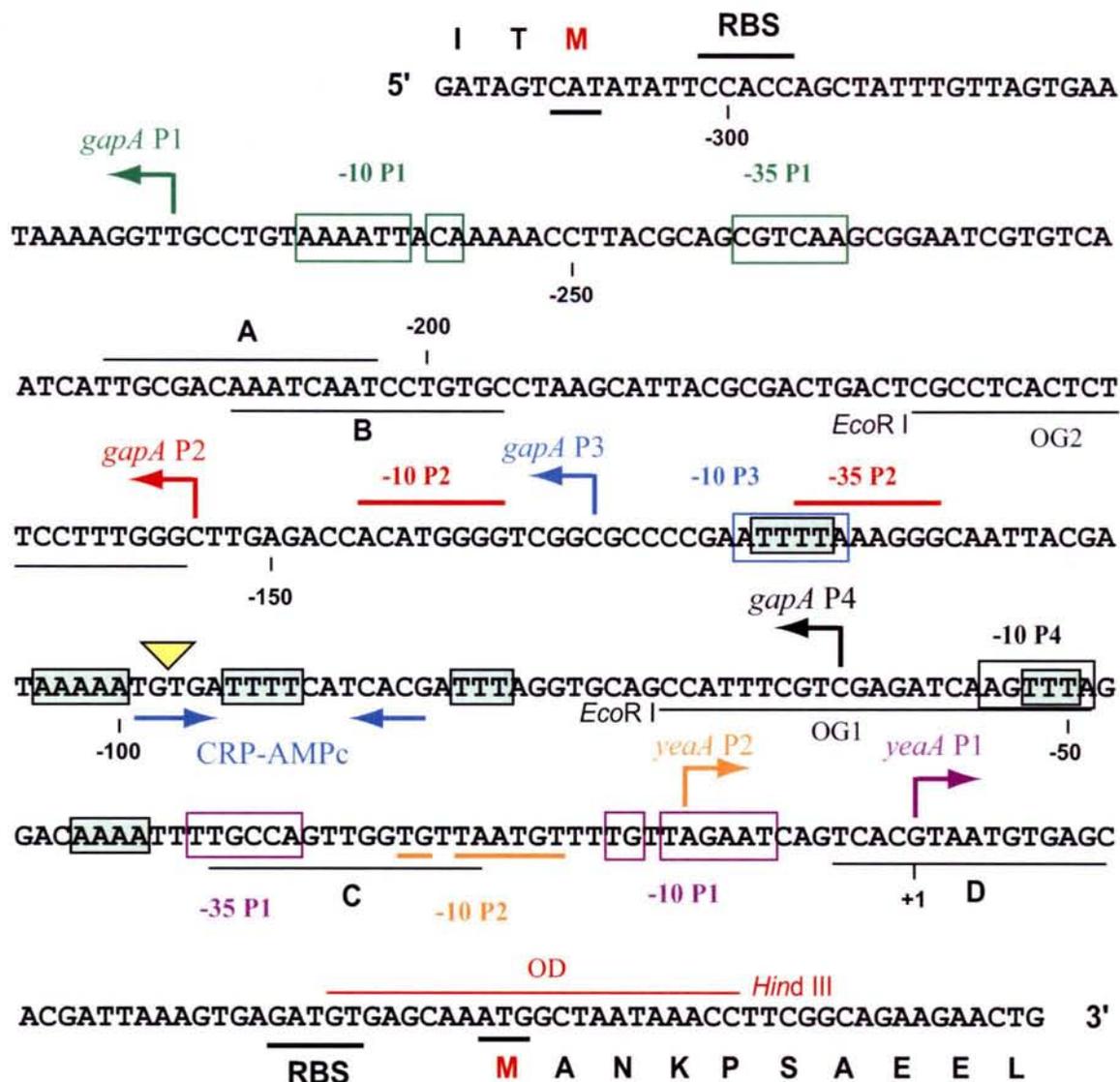


Figure 55 : Organisation génétique de la région inter-ORF *gapA-yeaA* d'*E. coli*.

Les 2 sites d'initiation de la transcription du gène *yeaA* sont indiqués par des flèches de couleurs différentes : *yeaA* P1 en violet et *yeaA* P2 en orange. La boîte -10 étendue et la boîte -35 du promoteur *yeaA* P1 sont encadrées, la boîte -10 étendue du promoteur *yeaA* P2 est soulignée. Les 4 sites potentiels de fixation de la protéine FIS A, B, C et D sont soulignés par des traits noirs (le 5^{ème} site potentiel présent au niveau de la séquence codante du gène *yeaA* n'est pas indiqué). Le centre d'une forte courbure de l'ADN identifiée par B. Charpentier (Charpentier, 1994) est matérialisé par un triangle jaune, et les alignements de résidus T ou A qui pourraient permettre la génération de cette courbure sont encadrés en vert. Le site potentiel de fixation du complexe CRP-AMPc (Charpentier et Branlant, 1994) est indiqué par deux flèches convergentes. Les 4 sites d'initiation de la transcription du gène *gapA* sont représentés par des flèches de couleurs différentes. Les boîtes -10 et -35 des différents promoteurs sont encadrées ou soulignées. Les couples d'oligonucléotides (OD avec OG1 ou OG2) utilisés pour amplifier un fragment d'ADN contenant les promoteurs *yeaA* P1 et *yeaA* P2 et le cloner dans le plasmide pRLG sont indiqués, ainsi que les sites de restriction utilisés pour ce clonage. RBS : site de fixation du ribosome.

transcrits *gapA* P1 dans la souche ne produisant pas la protéine Fis. Afin de nous affranchir du problème de nombre de copies de plasmide, nous avons répété ce type d'expérience avec des souches isogéniques C600 et C600*fis* non transformées. Le taux de transcrits produits à partir de l'ADN chromosomique était estimé par élongation d'amorce. Nos données confirment que le taux de transcrit est légèrement diminué en absence de production de la protéine Fis (Fig. 56 A). Pour vérifier l'implication de la protéine Fis dans ce changement de taux d'expression, nous avons construit par mutagenèse dirigée un plasmide pBS::*EcogapA* contenant un promoteur *gapA* P1 variant dans lequel le site potentiel de fixation de la protéine Fis centré à la position -62 (site A) était muté (variant mt-Fis, Fig. 56 B). Dans un premier temps, nous avons comparé le taux de transcrits initiés aux promoteurs *gapA* P1 sauvage et muté *in vivo* (Fig. 56 C). Pour cela, comme dans l'article n°1, nous avons utilisé le taux de transcrit *yeaA* P1 comme référence. Sur la base de cette référence, nous avons estimé que le taux de transcrits *gapA* P1 présent dans la souche exprimant le gène variant mt-Fis correspond à environ 85% de celui présent dans le cas du gène *gapA* sauvage, ce qui est compatible avec un faible effet activateur de la protéine Fis sur le promoteur *gapA* P1. Il est important de noter que les taux de transcrits *gapA* P2 et *gapA* P3 ne sont pas modifiés dans ces conditions (Données non montrées).

Seules des expériences réalisées *in vitro* en présence de l'holoenzyme $E\sigma^{70}$ et de la protéine Fis purifiée pouvaient permettre de savoir si la protéine Fis agit bien directement sur le promoteur *gapA* P1, ou si au contraire, l'effet activateur observé sur le promoteur *gapA* P1 est médié par l'activation de l'expression d'un facteur agissant lui-même directement sur le promoteur *gapA* P1 (Fig. 56 D1 et D2). Pour cela, nous avons incubé le plasmide pRLG::*gapAP1* en présence de l'holoenzyme $E\sigma^{70}$ et de protéine Fis, ceci en utilisant les conditions d'incubation décrites dans l'article n°1, à l'exception du fait que deux concentrations différentes d'ARN pol commerciale (0,1 et 0,05 U) ont été testées et que la protéine Fis était ajoutée, ceci à une concentration du même ordre de grandeur que celle utilisée par l'équipe de R. Gourse dans l'étude de l'action de la protéine Fis sur le promoteur *rrnB* P1 (200 nM) (Ross *et al.*, 1990; Aiyar *et al.*, 2002). La protéine Fis nous a été fournie par R. Wagner (Düsseldorf). Les résultats confirment que l'effet activateur de la protéine Fis sur l'initiation de la transcription au promoteur *gapA* P1 est faible, le facteur d'activation étant compris entre 20 et 30% (Fig. 56 D1). La réalisation de la même expérience de transcription *in vitro* avec le plasmide pRLG::*gapAP1*, dans lequel le site potentiel de fixation de la protéine Fis centré à la position -62 est muté, renforce l'idée que le faible effet activateur de la protéine Fis dépend bien de la présence de ce site, puisque après mutation de ce site,

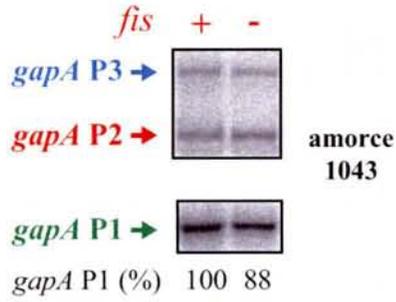
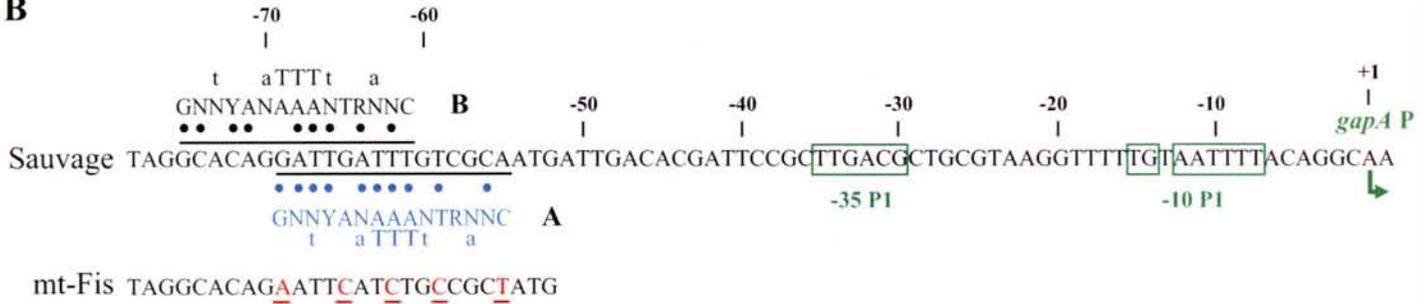
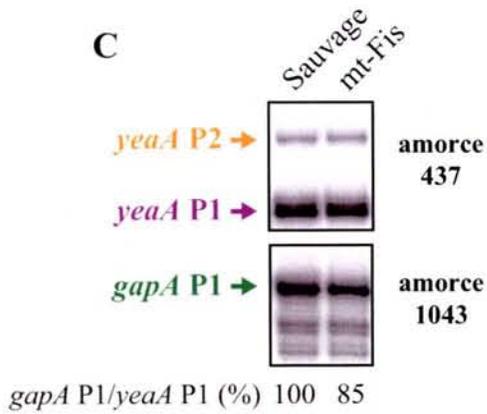
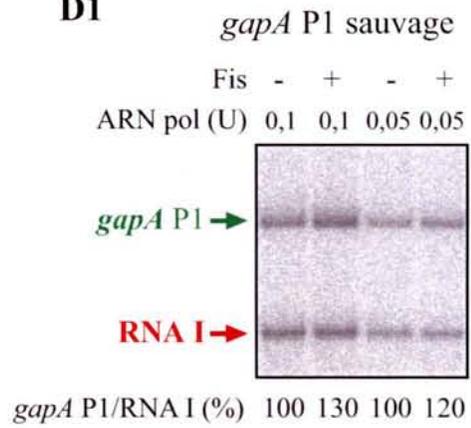
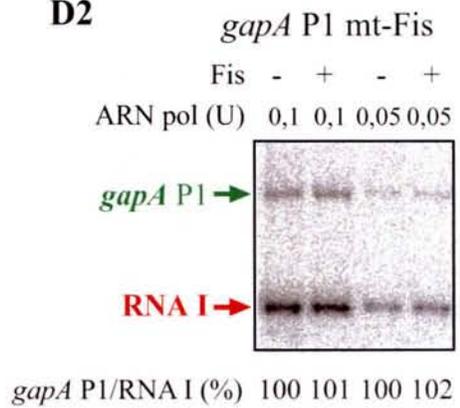
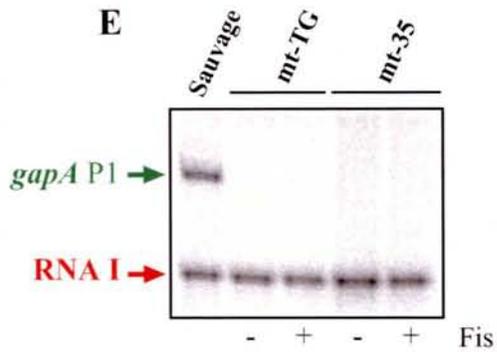
A**B****C****D1****D2****E**

Figure 56 : Effet de la protéine Fis sur l'initiation de la transcription au promoteur *gapA* P1.

A) Analyse des taux relatifs des ARNm *gapA* P1, *gapA* P2 et *gapA* P3 par élongation d'amorce par la transcriptase inverse dans les cellules des souches C600 et C600*fis*. Les bactéries ont été cultivées en milieu LB et un prélèvement de cellules a été effectué à une DO_{600nm} de 0,5. Les ARN totaux ont été extraits par la méthode au phénol chaud. Dans chaque essai, 20 µg d'ARN totaux ont été analysés par extension d'amorce par la transcriptase inverse. L'oligonucléotide amorce 1043 utilisé est complémentaire aux ARNm produits à partir des 3 promoteurs. Les ADNc obtenus ont été fractionnés en gel de polyacrylamide dénaturant à 5 %. La nature de l'ADNc *gapA* P1, *gapA* P2 et *gapA* P3 est indiquée à gauche du gel. La radioactivité contenue dans la bande correspondant aux ARNm *gapA* P1 a été quantifiée avec un PhosphorImager. La valeur obtenue pour le transcrit *gapA* P1 de la souche C600 a été arbitrairement prise comme référence (100 %). Celle obtenue pour la souche C600*fis* est exprimée en pourcentage par rapport à cette référence.

B) Représentation de la région promoteur *gapA* P1 avec les deux sites potentiels de fixation de la protéine Fis centrés aux positions -62 (site A) et -68 (site B) et les mutations générées dans cette région dans le variant mt-Fis. Les boîtes -10, -35 et le motif TG sont encadrés. Les sites potentiels de fixation de la protéine Fis sont soulignés et le consensus des séquences reconnues par la protéine Fis (Hengen *et al.*, 1997) est indiqué au-dessus du site A et au-dessous du site B. Lorsque deux bases alternatives sont présentes dans la séquence consensus, elles sont indiquées l'une sous l'autre. Les lettres en majuscule correspondent aux bases les plus fréquentes et celles en minuscule indiquent les préférences moins marquées (N indique A, C, G ou T). Les points noirs indiquent les bases identiques à celles présentes dans la séquence consensus.

Dans la séquence mt-Fis, les bases qui ont été mutées pour créer le promoteur variant sont indiquées en rouge.

C) Analyse par élongation d'amorce par la transcriptase inverse de l'effet de la mutation du site potentiel de fixation A de la protéine Fis sur le taux de transcrits *gapA* P1 présents *in vivo*. Les cellules *E. coli* XL1 blue ont été transformées par le plasmide pBS::*EcogapA* sauvage ou muté (mt-Fis). Les ARNm totaux extraits à une DO_{600nm} de 0,5 ont été analysés par extension des amorces 437 (transcrit *yeaA*, autoradiogramme du haut) et 1043 (transcrits *gapA* autoradiogramme du bas) marquées en 5' au ³²P. Pour chaque extension d'amorce, une même quantité d'ARN totaux (5 µg) a été utilisée. Les ADNc obtenus ont été fractionnés en gel de polyacrylamide dénaturant et la radioactivité présente dans les bandes d'ADNc a été mise en évidence et quantifiée par un PhosphorImager. Le rapport des radioactivités mesurées pour *gapA* P1 et *yeaA* P1, obtenu pour la souche transformée avec le gène sauvage a été pris arbitrairement comme référence (100 %). Le rapport obtenu pour la souche exprimant le gène muté a été exprimé en pourcentage de cette référence. Ces valeurs sont indiquées en-dessous des autoradiogrammes obtenus pour les oligonucléotides 437 et 1043 présentés dans la figure.

D) Expériences de transcription *in vitro* réalisées en présence de la protéine Fis purifiée, sur le promoteur *gapA* P1 sauvage (D1) ou muté dans le site Fis A (D2). Le plasmide pRLG::*gapA*P1 a servi de matrice et le protocole utilisé est identique à celui utilisé dans la publication n°1. Deux quantités différentes d'ARN pol commerciale (0,1 et 0,05 unités) ont été utilisées. La concentration de la protéine Fis purifiée était de 200 nM. Les transcrits *gapA* P1 et RNA I marqués par incorporation d'UTP [α -³²P] ont été mis en évidence par utilisation d'un PhosphorImager et ont été quantifiés. Le transcrit RNA I de 108 nt, produit par le plasmide pRLG, a été utilisé comme référence. Pour chaque concentration d'ARN pol, le rapport des taux de transcrits *gapA* P1/RNA I obtenus en absence de protéine Fis a été pris comme référence (100 %) et la valeur de ce rapport obtenue en présence de Fis a été exprimée en pourcentage de cette référence.

E) La protéine Fis ne permet pas d'activer la transcription *in vitro* à partir des promoteurs *gapA* P1 mutés au niveau du dinucléotide TG et au niveau de la boîte -35. Les variants mt-TG et mt-35 utilisés sont décrits dans la publication n°1. La protéine Fis purifiée a été ajoutée au mélange de transcription (+) à une concentration de 200 nM. Les transcrits *gapA* P1 et RNA I sont indiqués à gauche de l'autoradiogramme.

nous n'observons plus d'augmentation du taux de transcrits *gapA* P1 en présence de la protéine Fis (Fig. 56 D2). Du fait de l'architecture particulière du promoteur *gapA* P1 décrite dans l'article n°1, il était intéressant de comparer l'activité de la protéine Fis vis-à-vis du promoteur *gapA* P1 sauvage et de ces deux variants mt-TG et mt-35 (article n°1) dans lesquels le dinucléotide TG et la boîte -35 sont respectivement mutés. Les données présentées dans la figure 56 E montrent que même en présence de la protéine Fis, le taux de transcrits initiés à partir de ces deux promoteurs mutés reste très faible. En conclusion, *in vitro*, la protéine Fis a un léger effet activateur sur l'initiation de la transcription au promoteur *gapA* P1. Comme nous l'avons noté plus haut, le site A de fixation de la protéine Fis chevauche l'élément distal (-50 -60) protégé par l'ARN pol dans le complexe ouvert RPo. Il sera intéressant de savoir comment la fixation de la protéine Fis à cette position modifie l'interaction entre l'ADN et l'ARN pol. Il sera bien sûr aussi intéressant de réaliser le même type d'étude après mutation du site potentiel B seul et mutation simultanée des deux sites A et B.

II.2) Etude de l'effet de la protéine Fis sur l'activité des promoteurs *yeaA* P1 et *yeaA* P2.

Comme nous l'avons mentionné ci-dessus, les 3 autres sites potentiels de fixation de la protéine Fis sont situés en amont du promoteur *gapA* P4. Deux sites chevauchent les boîtes -10 et -35 potentielles des promoteurs *yeaA* P1 et *yeaA* P2, ainsi que le site (+1) d'initiation de la transcription du promoteur *yeaA* P1, le troisième étant situé très en amont au niveau de la séquence codante de la protéine MsrB (Fig. 45 et 55). On pouvait donc se demander si la protéine Fis limite l'utilisation des promoteurs *yeaA* P1 et *yeaA* P2. Pour répondre à cette question, nous avons utilisé les souches C600 et C600*fis* et comparé le taux de transcrits *yeaA* P1 et *yeaA* P2 dans ces deux souches. Pour cela, nous avons utilisé la technique d'extension d'amorce et l'oligonucléotide 437 décrit dans l'article n°1 (Fig. 57 B). Etant donné que la protéine Fis est exprimée fortement en début de phase exponentielle de croissance et que sa concentration dans les cellules diminue, pour être quasiment nulle en phase stationnaire de croissance (Ball *et al.*, 1992), l'analyse a été réalisée sur des échantillons prélevés à des DO_{600nm} de 0,2 (1h30 de culture, début de la phase exponentielle), 0,5 (2h de culture, milieu de phase exponentielle) et 2 (4h de culture, fin de phase exponentielle). Nous observons que le taux de transcrits *yeaA* P1 est supérieur dans la souche C600*fis* par rapport à celui dans la souche C600 (Fig. 57 B) (facteur de 1,4), ceci pour les 3 stades de la croissance que nous

avons étudiés. Les données sont en accord avec un rôle répresseur de la protéine Fis sur l'initiation de la transcription au promoteur *yeaA* P1. Il est à noter que le taux de transcrit initié au promoteur *yeaA* P2 est très faible dans ces conditions de croissance, aussi bien en présence qu'en absence d'expression de la protéine Fis et que des bandes supplémentaires apparaissent en fin de phase exponentielle (Fig. 57 B). Alors que la comparaison globale des données sur les deux souches est en accord avec un effet inhibiteur de la protéine Fis, la faible augmentation du taux de transcrits *yeaA* P1 en fonction du temps de culture l'est moins. En effet, la concentration en protéine Fis est supposée diminuer au cours de la phase exponentielle de croissance pour être nulle en phase stationnaire de croissance. Une hypothèse serait que l'affinité de la protéine Fis pour ces sites de fixation soit suffisamment importante pour que la quantité de cette protéine encore présente dans les cellules dont la DO_{600nm} a atteint la valeur de 2, soit encore suffisante pour permettre l'inhibition de l'initiation de la transcription à partir du promoteur *yeaA* P1. Pour compléter ces expériences, il sera nécessaire de réaliser à nouveau une expérience avec des prélèvements à différents stades de la phase stationnaire. Le taux de protéine Fis contenue dans les cellules collectées à ces différents stades de croissance devra être suivi par emploi d'un anticorps. Une autre possibilité était que Fis n'agisse pas directement sur la transcription du gène *yeaA* mais par l'intermédiaire d'un facteur agissant directement sur les promoteurs du gène *yeaA*. Pour tester si l'effet de la protéine Fis était direct ou indirect, nous avons inséré un segment d'ADN contenant les deux promoteurs *yeaA* P1 et *yeaA* P2 dans le plasmide pRLG770. Cette construction appelée pRLG::*yeaA*PL contient les régions des deux promoteurs, la région aval allant jusqu'au codon initiateur de la traduction de l'ORF *yeaA* et la région qui permet la fixation du complexe CRP-AMPc (position -174 à +49 par rapport au +1) (Fig. 55). Les oligonucléotides OD et OG2 utilisés pour réaliser cette construction sont représentés dans la figure 55.

Dans un premier temps, nous avons réalisé une expérience de transcription *in vitro* en utilisant comme matrice le plasmide pRLG::*yeaA*PL en présence de concentrations croissantes de protéine Fis purifiée. Les résultats sont présentés dans la figure 57 C. Le taux de transcrits initiés au promoteur *yeaA* P1 est fortement diminué lorsque la concentration de la protéine Fis augmente. L'évaluation du rapport des concentrations des transcrits *yeaA* P1/RNA I montre qu'une concentration en protéine Fis d'environ 40 nM conduit à une diminution d'environ 50 % du taux de transcrits initiés au promoteur *yeaA* P1. A titre de comparaison, la concentration en protéine Fis requise pour limiter la transcription *in vitro* à partir des promoteurs *gyrA* et *gyrB* est de l'ordre de 50 nM (Schneider *et al.*, 1999). Ces données suggèrent qu'une faible

concentration en protéine Fis est suffisante pour avoir un effet inhibiteur sur le promoteur *yeaA* P1. De plus, ces données suggèrent que l'affinité de la protéine Fis pour la région promoteur du gène *yeaA* est du même ordre de grandeur que celle pour les régions promoteur d'autres gènes pour lesquels l'effet inhibiteur de la protéine Fis a été démontré. Le gène *yeaA* ferait donc partie des gènes régulés négativement par la protéine Fis. Il est à noter qu'*in vitro*, aucune initiation de la transcription n'a lieu à partir du promoteur *yeaA* P2. Sur la base de cet effet très marqué de la protéine Fis *in vitro*, comment peut-on interpréter les résultats obtenus *in vivo* (Fig. 57 B) ?

La faible concentration en protéine Fis nécessaire pour qu'il y ait inhibition *in vitro* pourrait expliquer que la concentration en protéine Fis présente à une DO_{600nm} de 2 soit encore suffisante pour qu'il y ait une inhibition encore marquée de la transcription au promoteur *yeaA* P1. Par ailleurs, l'utilisation du promoteur *yeaA* P1 dans des cellules en début ou en milieu de phase exponentielle de croissance, où la concentration de la protéine Fis est élevée, peut s'expliquer par l'existence de une ou plusieurs molécules activatrices contrecarrant l'effet de Fis *in vivo*. Ceci pourrait permettre une production minimale de protéine MsrB en phase exponentielle de croissance.

On peut imaginer que l'effet répresseur de Fis sur la transcription du gène *yeaA* *in vivo* conduise à limiter l'expression de la protéine MsrB en phase exponentielle de croissance, et que son expression soit renforcée en phase stationnaire de croissance du fait de l'absence de la protéine Fis. L'existence de cette régulation du taux de production de la protéine MsrB en fonction de la phase de croissance pourrait s'expliquer par le fait qu'en phase stationnaire, les cellules sont soumises à un grand nombre de stress, en particulier de stress oxydant d'où la nécessité d'exprimer plus de protéine MsrB. Un grand nombre d'autres protéines de résistance au stress sont d'ailleurs aussi exprimées au cours de cette phase (cf. introduction). L'équipe de F. Barras (Marseille) a obtenu des résultats qui confortent l'hypothèse d'une augmentation de la production de la protéine MsrB en phase stationnaire. En effet, cette équipe a montré une augmentation du taux de production de β -glucuronidase en phase stationnaire de croissance dans des cellules d'*E. coli* renfermant une copie unique au niveau chromosomique d'une construction du type fusion transcriptionnelle réalisée avec la région promoteur du gène *yeaA* et l'ORF *uidA* codant la β -glucuronidase (Ezraty, 2004). De plus, cette même équipe a montré, en utilisant une souche mutée au niveau du gène *rpoS* codant le facteur σ^S , que cette augmentation n'est pas dépendante du facteur σ spécifique de la phase

stationnaire (Ezraty, 2004). Nous proposons donc un modèle de régulation du taux d'expression du gène *yeaA* dans lequel la protéine Fis joue le rôle de répresseur pour limiter l'expression en phase exponentielle de croissance.

III.) Le promoteur *yeaA* P1 fait partie de la famille des promoteurs étendus.

Etant donné que le promoteur *yeaA* P1 était utilisé efficacement *in vitro*, nous nous sommes intéressés à l'architecture de ce promoteur. L'examen de sa séquence montre qu'une boîte -10 potentielle est présente 8 résidus en amont du site d'initiation des transcrits *yeaA* P1 que nous avons identifié par élongation d'amorce. De plus, 5 des 6 paires de bases de cet élément -10 potentiel correspondent au consensus des promoteurs reconnus par $E\sigma^{70}$ (TAGAAT). Bien que dans ce cas, contrairement à la situation observée pour le promoteur *gapA* P1, la séquence de la boîte -10 soit très similaire à la séquence consensus des boîtes -10, comme pour *gapA* P1, un dinucléotide TG est présent une paire de base en amont de la boîte -10. Ainsi, le promoteur *yeaA* P1 pouvait aussi être un promoteur étendu. Enfin, une séquence très similaire à la séquence consensus des éléments -35 des promoteurs reconnus par $E\sigma^{70}$ (TTGCCA ; 5 bases sur 6 identiques au consensus) se situe 19 pb en amont de la boîte -10. Cette distance est élevée par rapport à la distance optimale de 17 pb.

Comme lors de l'étude du promoteur *gapA* P1, nous avons vérifié l'importance du dinucléotide TG pour l'efficacité transcriptionnelle du promoteur *yeaA* P1 par mutagenèse dirigée. La mutagenèse a été réalisée sur le plasmide pBS::*EcogapA*. Nous avons muté le résidu G du dinucléotide TG en résidu C (variant mt-TG). Nous avons aussi testé l'effet du remplacement du résidu T à l'extrémité 3' de la boîte -10 par un résidu C (variant mt-10) (Fig. 58 A). Enfin, nous avons testé l'importance de la boîte -35 potentielle pour l'efficacité de ce promoteur en créant un variant (mt-35), où les 3 résidus en 5' de cette séquence sont remplacés par des résidus complémentaires (Fig. 58 A). L'efficacité des promoteurs mutés a été testée *in vivo* en transformant des cellules d'*E. coli* XL1 blue par les plasmides variants et en analysant le taux des transcrits *yeaA* P1 par élongation d'amorce à la transcriptase inverse (Fig. 58 B). Le taux de transcrits initiés au promoteur *gapA* P1, qui était supposé être inchangé par les mutations réalisées, nous a servi de référence. Le rapport des taux de transcrits *yeaA* P1/*gapA* P1, observé pour le promoteur sauvage, a été utilisé comme référence (100 %).

Les résultats montrent que la mutation du dinucléotide TG diminue très fortement le taux de transcrits initiés au promoteur *yeaA* P1 (facteur 16) (Fig. 58 B). Comme la mutation de la paire de bases située à l'extrémité 3' de la boîte -10 diminue aussi l'efficacité du promoteur d'un facteur 50, nous en concluons que le promoteur *yeaA* P1 utilise cette boîte -10 TAGAAT et qu'il appartient à la famille des promoteurs étendus. La mutation de la boîte -35 potentielle ne diminue l'efficacité de la transcription que d'un facteur 2 (Fig. 58 B). Nous en concluons donc que contrairement à ce que nous avons observé pour le promoteur *gapA* P1, la boîte -35 potentielle n'est pas essentielle à l'efficacité du promoteur *yeaA* P1. Nous avons ensuite vérifié les résultats obtenus *in vivo* par des expériences de transcription *in vitro*. Pour cela, nous avons construit le plasmide pRLG::*yeaAPc* qui contient un fragment d'ADN amplifié par PCR à partir de la région *yeaA-gapA* avec les oligonucléotides OD et OG1 (Fig. 55). Le fragment amplifié comporte les deux promoteurs et la région située en aval, allant jusqu'au codon initiateur de la traduction de la protéine MsrB (position -71 à +49 par rapport au site +1 *yeaA* P1). Des plasmides pRLG::*yeaAPc* mt-10, mt-TG et mt-35 dans lesquels nous avons, respectivement, introduit les mutations en 3' de la boîte -10, du dinucléotide TG et de la boîte -35 ont aussi été produits (Fig. 58 A). La présence du dinucléotide TG, ainsi que l'intégrité de la boîte -10 s'avèrent essentielles pour qu'il y ait transcription *in vitro*, alors que les mutations dans la boîte -35 ne réduisent l'efficacité de transcription *in vitro* que de 20 % (Fig. 58 C). Ces données confirment celles obtenues *in vivo*.

En conclusion, le promoteur *yeaA* P1 est un promoteur ayant des caractéristiques intermédiaires entre celles du promoteur *gapA* P1, qui nécessite une boîte -35 pour être actif, et celle du promoteur P_{RE}* pour lequel la séquence de la boîte -35 n'a que très peu d'influence sur l'efficacité du promoteur (Keilty et Rosenberg, 1987). En effet, pour le promoteur *yeaA* P1, bien que la boîte -35 ne soit pas essentielle, elle augmente l'efficacité de transcription d'un facteur 2. Des résultats similaires ont été obtenus par d'autres auteurs pour le promoteur *galP1*, pour lequel l'insertion d'une séquence -35 canonique à la place de la séquence sauvage augmente le taux d'initiation de la transcription (Chan et Busby, 1989).

IV.) Rôle de la forte courbure et du site de fixation du complexe CRP-AMPc présents dans la région intergénique *gapA-yeaA* sur l'efficacité des promoteurs *yeaA* P1 et *yeaA* P2.

Afin de tester l'importance possible de la région d'ADN comportant la forte courbure centrée en position -97 sur l'efficacité d'initiation de la transcription par le promoteur *yeaA* P1, nous avons comparé l'efficacité de transcription *in vitro* à partir du promoteur *yeaA* P1 contenu dans le plasmide pRLG::*yeaAPc*, qui ne contient pas cette région courbée et le plasmide pRLG::*yeaAPL*, qui la contient (Fig. 58 C). On constate que l'efficacité du promoteur *yeaA* P1 est renforcée dans la construction pRLG::*yeaAPL* par rapport à la construction pRLG::*yeaAPc* (facteur d'environ 1,7). Ceci suggère que la forte courbure qui, d'après les données antérieures de B. Charpentier est centrée en position -97 par rapport au site d'initiation de la transcription du promoteur *yeaA* P1, influence l'efficacité de ce promoteur.

Comme nos données de transcription *in vitro* montraient que le promoteur *yeaA* P1 est plus actif en présence de la forte courbure amont, la question était de savoir s'il était aussi activé par la fixation du complexe CRP-AMPc, dont le site de fixation chevauche la région courbée. Par emploi d'une souche *cya*⁻, B. Charpentier avait montré l'importance du complexe CRP-AMPc pour l'activité du promoteur *gapA* P3 (Charpentier et Branlant, 1994). Le site de fixation de ce complexe étant situé 41 pb en amont du site d'initiation de la transcription de *gapA* P3, on pouvait imaginer que le complexe CRP-AMPc joue un rôle d'activateur de classe II. La mutation de ce site de fixation affectait aussi fortement l'activité du promoteur *gapA* P3 mais cette mutation touchait en même temps la courbure. De ce fait, les expériences de mutagenèse n'avaient pas permis de distinguer les rôles respectifs de la courbure et du site de fixation du complexe CRP-AMPc dans l'activité des promoteurs *gapA* P3 et *gapA* P2.

Afin de distinguer les effets respectifs de ces deux éléments sur les 4 promoteurs *gapA* P2, *gapA* P3, *yeaA* P1 et *yeaA* P2, nous avons utilisé les variants mt-421, mt-471 et mt-474, que B. Charpentier avait construit (Fig. 59 A) (Charpentier, 1994) et pour chacun de ces variants, nous avons évalué l'amplitude de la courbure par la technique d'électrophorèse en gel non dénaturant décrite dans l'article n°1 (Fig. 59).

Dans le variant mt-421, la séquence mutée correspond à un des éléments de séquence correspondant au motif de reconnaissance du complexe CRP-AMPc (Fig. 59 A). Par ailleurs, les substitutions de bases réalisées affectent la série de résidus A (positions -100 à -104) en

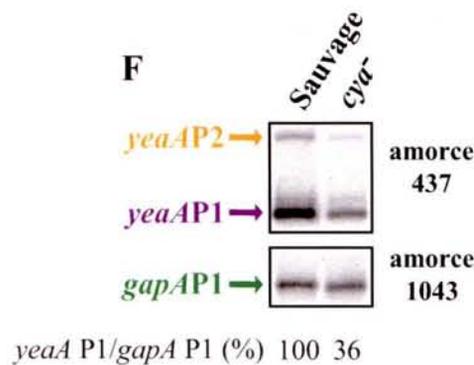
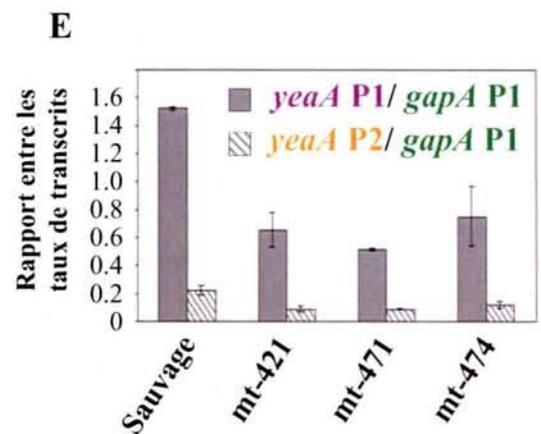
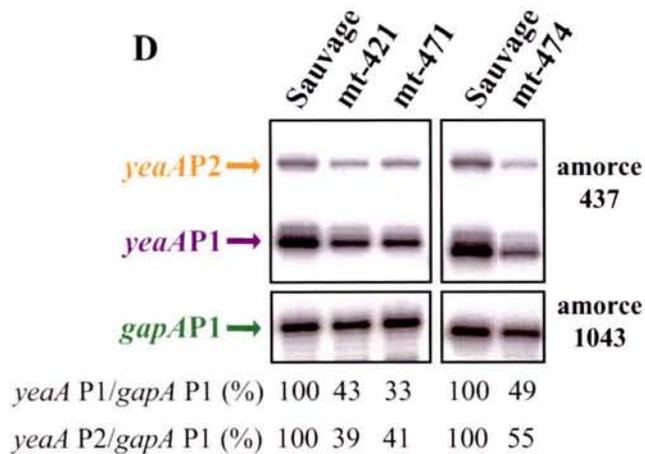
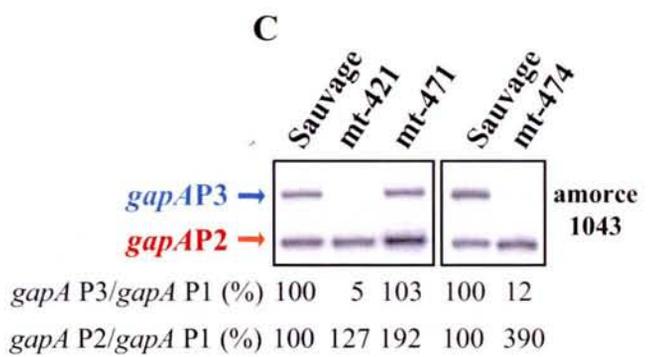
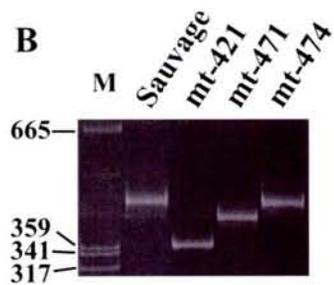
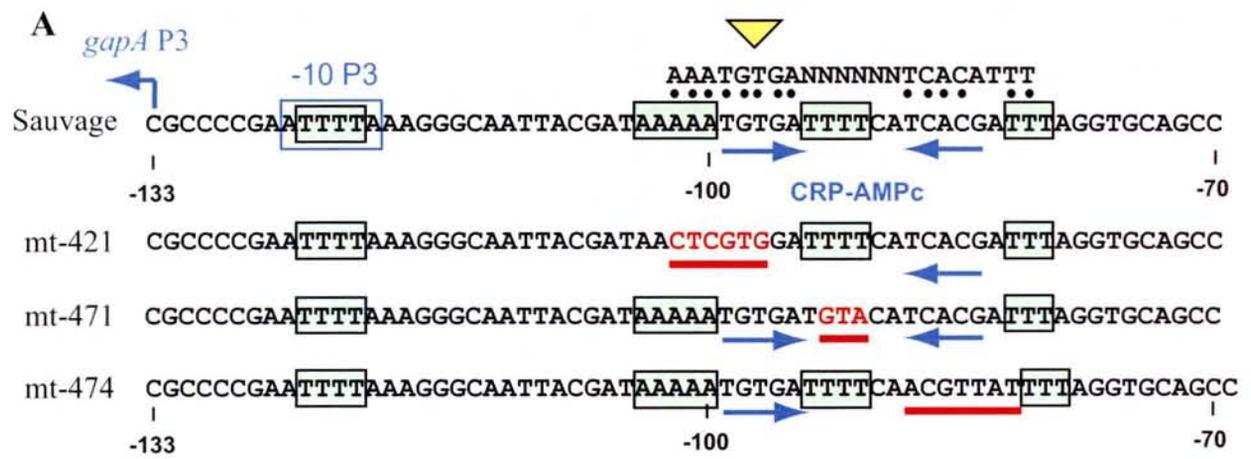


Figure 59 : Effet de la forte courbure présente dans la région intergénique *gapA-yeaA* sur l'initiation de la transcription aux promoteurs *yeaA* P1 et *yeaA* P2.

A) Séquence des variants créés par B. Charpentier. Le site d'initiation de la transcription P3 est indiqué par une flèche bleue. La boîte -10 du promoteur *gapA* P3 est encadrée. Le site de fixation du complexe CRP-AMPC est indiqué par des flèches inversées. La séquence consensus du site de fixation de ce complexe est indiquée au-dessus de la séquence et les bases identiques à la séquence consensus sont indiquées par un point noir. Les mutations sont soulignées en rouge et le centre de la courbure est indiquée par un triangle jaune. Les alignements de résidus T et A qui seraient impliqués dans la génération de la courbure sont encadrés en vert.

B) Electrophorèse en gel non dénaturant des fragments amplifiés par PCR (cf. experimental procedure article n°1) correspondant aux séquences des variants présentés dans le panneau A). Un marqueur de taille M a été déposé dans la première piste.

C) Analyse du taux de transcrits initiés aux promoteurs *gapA* P2 et *gapA* P3 pour les variants mt-421, mt-471 et mt-474 par élongation d'amorce à la transcriptase inverse. Les cellules *E. coli* XL1 blue ont été transformées par les plasmides pBS::*EcogapA* sauvage ou mutés. Les ADNc correspondant aux transcrits initiés aux promoteurs *gapA* P2 et *gapA* P3 sont indiqués. Les rapports *gapA* P3/*gapA* P1 et *gapA* P2/*gapA* P1 sont indiqués en-dessous des autoradiogrammes. La valeur du rapport obtenu pour le sauvage est pris arbitrairement comme référence (100 %).

D) Analyse par élongation d'amorce par la transcriptase inverse du taux de transcrits initiés aux promoteurs *yeaA* P1 et *yeaA* P2. Le taux de transcrits *gapA* P1 a été utilisé comme référence. Les ADNc correspondant aux transcrits initiés aux promoteurs *yeaA* P1 et *yeaA* P2 pour les variants mt-421, mt-471 et mt-474 sont indiqués.

E) Histogramme représentant l'effet des mutations sur le taux de transcrits initiés aux promoteurs *yeaA* P1 et *yeaA* P2. Le taux de transcrits initiés au promoteur *gapA* P1 a été utilisé comme référence du nombre de copie du plasmide. Les bandes en B) ont été quantifiées par phosphorimager (unités Molecular Dynamics) et les rapports *yeaA* P1/*gapA* P1 et *yeaA* P2/*gapA* P1 ont été calculés. Les erreurs standard sont indiquées par des barres noires.

F) Analyse du taux de transcrits initiés aux promoteurs *yeaA* P1 et *yeaA* P2 dans une souche *cya*⁻ par élongation d'amorce à la transcriptase inverse. Le taux de transcrits *gapA* P1 a été utilisé comme référence.

phase avec la série de T (positions -91 à -94), qui est vraisemblablement impliquée dans la génération de la courbure. Comme attendu, nous observons une très forte diminution de la courbure de l'ADN, la mobilité électrophorétique du fragment étant fortement augmentée après la mutation (Fig. 59 B). Donc, pour ce variant, à la fois la courbure et le site de fixation du complexe CRP-AMPc sont affectés.

Pour le variant mt-471, les éléments de séquence requis pour la fixation du complexe CRP-AMPc n'étaient pas modifiés, mais un des enchaînements de résidus T supposé participer à la génération de la courbure, qui est situé entre les deux séquences répétées inversées requises pour fixer le complexe CRP-AMPc a été muté (Fig. 59 A). En accord avec ces caractéristiques, l'amplitude de la courbure est réduite pour ce variant (Fig. 59 B). On peut donc supposer que ce variant présente une plus faible courbure mais conserve sa capacité à fixer le complexe CRP-AMPc. Enfin, pour le variant mt-474, les enchaînements de résidus A et T en phase avec le pas de l'hélice sont conservés. Comme attendu, la courbure est peu affectée par la mutation (Fig. 59 B) mais un des éléments de séquence requis pour la fixation du complexe CRP-AMPc est fortement altéré. On s'attend donc à avoir pour ce variant une forte courbure mais une incapacité à fixer le complexe CRP-AMPc.

Pour chacun de ces variants, nous avons testé l'efficacité d'initiation de la transcription aux promoteurs *gapA* P2 et *gapA* P3 *in vivo*. La perte de la courbure et de la capacité à fixer le complexe CRP-AMPc dans le variant mt-421 abolit presque totalement l'activité du promoteur *gapA* P3, alors que l'activité du promoteur *gapA* P2 est très faiblement augmentée (Fig. 59 C). Pour le variant mt-474 qui conserve une forte courbure et ne doit pas pouvoir fixer le complexe CRP-AMPc, l'activité du promoteur *gapA* P3 est fortement diminuée, alors que l'activité du promoteur *gapA* P2 est augmentée d'un facteur 4 (Fig. 59 C). Enfin, pour le variant mt-471 dans lequel le site de fixation du complexe CRP-AMPc doit être préservé, alors que le degré de courbure est réduit, l'activité du promoteur *gapA* P3 est similaire à celle obtenue pour le plasmide sauvage. Par contre, une augmentation d'un facteur 2 de l'activité du promoteur *gapA* P2 est observée.

Ces données montrent que les enchaînements de résidus T et A que nous avons mutés sont en partie responsables de la génération de la forte courbure. D'autres enchaînements de résidus T ou A doivent être impliqués dans la génération de cette courbure du fait de son fort degré. Il existe plusieurs enchaînements de résidus A ou T en phase avec les deux blocs situés respectivement entre les positions -100 à -104 et -91 à -94 (Fig. 55 et 59) qui sont, comme nous venons de voir, impliqués dans la courbure. Ces enchaînements indiqués dans la figure

55 pourraient contribuer à la courbure. Il est intéressant de noter que l'un de ces enchaînements (4 résidus T) est présent dans la boîte -10 du promoteur *gapA* P3. Pour tester leur implication, de nouveaux mutants devront être produits.

Sur la base de l'ensemble de nos données, il semble que l'existence d'une courbure ne soit pas un paramètre important pour l'activité du promoteur *gapA* P3. Par contre, nous confirmons bien sa dépendance vis-à-vis du complexe CRP-AMPc. Les données ne permettent pas de trouver un lien direct entre l'activité du promoteur *gapA* P2 et la présence de la courbure. Un seul point ressort clairement dans l'état actuel des données, l'activité du promoteur *gapA* P2 ne dépend pas de la fixation du complexe CRP-AMPc. Elle pourrait même être inhibée par cette fixation. La génération de nouveaux variants sera nécessaire pour essayer d'expliquer les fortes augmentations de l'activité du promoteur *gapA* P2 observées pour les variants mt-471 et mt-474.

Nous avons ensuite testé l'effet des 3 mêmes mutations sur l'efficacité de l'initiation de la transcription aux promoteurs *yeaA* P1 et *yeaA* P2 (Fig. 59 D et E). Il apparaît que quelque soit la mutation réalisée dans cette séquence entre les positions -82 à -102, le taux de transcrits initiés aux promoteurs *yeaA* P1 et *yeaA* P2 est diminué d'un facteur compris entre 2 et 3, ceci indépendamment de l'effet de la mutation sur la courbure ou sur le site de fixation du complexe CRP-AMPc (Fig. 59 E). En plus, la baisse d'activité n'est pas proportionnelle au niveau d'altération de la courbure. Soit la courbure et le site de fixation du complexe CRP-AMPc ont des effets conjugués sur l'activité des promoteurs *yeaA* P1 et *yeaA* P2, et/ou d'autres facteurs protéiques se lient à cette région et interviennent dans l'activité des deux promoteurs.

Pour tester l'hypothèse d'un rôle du complexe CRP-AMPc sur les deux promoteurs *yeaA* P1 et *yeaA* P2, nous avons comparé les taux de transcrits *yeaA* P1 et *yeaA* P2 dans la souche TP2006 *cya*⁻ (ne produisant pas d'AMPc) et dans la souche isogénique TP 2503 produisant l'AMPc (Fig. 59 F). Le taux de transcrits initiés au promoteur *yeaA* P1 est diminué d'un facteur 3 dans la souche *cya*⁻ et l'ARNm *yeaA* P2 est quasiment indétectable dans cette souche. La similarité des effets observés en mutant le site CRP-AMPc et en ne produisant pas d'AMPc suggère fortement que l'effet du complexe CRP-AMPc sur l'expression du gène *yeaA* est direct. Nous proposons donc que la fixation du complexe CRP-AMPc, situé entre les positions -84 et -99 par rapport au site +1 *yeaA* P1, participe fortement à l'activation de la transcription du gène *yeaA*. Le site de fixation de ce complexe est centré à la position -79,5 par rapport au site +1 du promoteur P2 et -91,5 par rapport au promoteur P1. Ces distances

sont compatibles avec un effet du complexe CRP-AMPc en tant qu'activateur de classe I. La concentration en AMPc augmentant en fin de phase exponentielle du fait de la diminution de la vitesse de croissance (Kuo *et al.*, 2003), on peut donc penser que l'activation par le complexe CRP-AMPc puisse contrecarrer l'effet inhibiteur de la protéine Fis dans cette phase de croissance. La fixation du complexe CRP-AMPc au site -84 à -99 aurait le double effet d'activer le promoteur *gapA* P3 et les deux promoteurs *yeaA* P1 et *yeaA* P2. Il est difficile de tirer des conclusions sur l'effet de la courbure, étant donné que les 3 promoteurs variants produits, qui ont des degrés de courbure différents, ont des activités peu différentes les unes des autres.

Une explication possible des données pourrait être que la formation des complexes RPo au niveau des promoteurs *yeaA* P1 et *yeaA* P2 soit facilitée par la fixation du complexe CRP-AMPc sur le site -84 à -99, et que l'interaction entre ARN pol et le complexe CRP-AMPc ne soit optimale que pour un degré de courbure bien défini. Altérer la courbure ou le site CRP-AMPc, ou altérer les deux aboutirait aux mêmes effets négatifs que ceux observés en absence d'AMPc.

Une autre possibilité est qu'*in vivo*, en plus du complexe CRP-AMPc, d'autres facteurs protéiques comme par exemple les protéines histone-like H-NS et HU, qui se fixent préférentiellement sur des régions d'ADN courbées (Azam et Ishihama, 1999), interviennent dans la régulation de l'activité des promoteurs *yeaA* P1 et *yeaA* P2 en se fixant sur le segment -100 à -70. Les sites de fixation de la protéine Fis que nous avons décrits ci-dessus sont localisés à 68 et 101 pb du centre estimé de la courbure. Il sera important de tester si l'action inhibitrice observée de la protéine Fis est contrecarrée ou au contraire augmentée par la présence de la courbure centrée en position -97. Pour cela, des essais de transcription *in vitro* devront être réalisés avec les variants mt-421, mt-474 et mt-471 en présence ou en absence de la protéine Fis et du complexe CRP-AMPc. Il sera aussi intéressant d'utiliser dans ce type d'expérience, des mutants de l'ARN pol dans lesquels le domaine α -CTD est délété.

V.) Etude sur l'implication possible de l'ARN régulateur RyhB dans la régulation de la traduction du transcrit *yeaA*.

L'équipe de F. Barras à Marseille qui a trouvé que la protéine codée par le gène *yeaA* possède une activité MsrB (Grimaud *et al.*, 2001), a également observé, en utilisant des fusions entre

le gène *yeaA* et le gène codant la β -glucuronidase, que la présence d'ions ferriques (Fe^{2+}) pourrait contrôler le taux d'expression du gène *yeaA* (Ezraty, 2004). Il y aurait augmentation de l'expression de ce gène en présence d'ion Fe^{2+} et cette augmentation d'expression pourrait être dépendante de la présence du facteur de transcription Fur ((Ezraty, 2004) ; F. Barras, communication personnelle). Comme un certain nombre de gènes dont l'expression est augmentée en présence d'ions Fe^{2+} sont régulés au niveau post-transcriptionnel par un petit ARN non codant RyhB, nous avons essayé de savoir si en plus des régulations transcriptionnelles décrites ci-dessus, le gène *yeaA* pouvait être régulé au niveau post-transcriptionnel par l'ARN RyhB. Je commencerai donc par présenter le système RyhB.

V.1) Présentation de l'ARN antisens RyhB et de son implication dans l'homéostasie du fer.

Chez *E. coli* et chez beaucoup d'autres bactéries, les systèmes assurant la régulation de la concentration en ions Fe^{2+} dans la cellule sont très importants pour la survie des cellules. En effet, l'ion Fe^{2+} est un nutriment essentiel à la croissance, mais lorsqu'il est présent en excès, il peut causer des dommages à la cellule, notamment par génération d'entités oxygénées réactives (Imlay, 2003). Chez *E. coli*, la plupart des régulations de l'expression des gènes, qui sont liées à la présence d'ions Fe^{2+} , dépendent du facteur de transcription Fur (Ferric Uptake Regulator protein), qui est un régulateur transcriptionnel (pour revue : Hantke, 2001; Andrews *et al.*, 2003). Lorsque les ions Fe^{2+} sont abondants, la protéine Fur qui est activée par fixation de Fe^{2+} , peut se lier à des éléments de régulation de l'expression des gènes présents sur l'ADN. Ceci permet la répression de gènes qui codent des protéines impliquées dans le transport du fer à l'intérieur de la cellule (Fig. 60) (Andrews *et al.*, 2003). Si le taux d'ions Fe^{2+} diminue, seulement une partie de la protéine Fur est liée aux ions Fe^{2+} , ses capacités de fixation à l'ADN diminuent et la répression des gènes contrôlés par ce facteur de transcription est diminuée, voire levée. En plus des gènes réprimés par la protéine Fur, il existe des gènes régulés positivement par cette protéine en présence d'ions Fe^{2+} . Dans ce cas, le mécanisme est indirect, il passe en général par l'action du petit ARN RyhB, dont la transcription est régulée négativement par la protéine Fur (Fig. 60). Ces gènes codent des protéines non essentielles de stockage du fer, en particulier, des protéines contenant un centre fer-soufre. L'ARN régulateur non codant (ARNnc) RyhB de 90 nucléotides a d'abord été mis en évidence par des études bioinformatiques, réalisées sur le génome d'*E. coli* (Argaman *et al.*, 2001; Wassarman *et al.*, 2001). L'ARN RyhB peut s'hybrider spécifiquement aux ARNm codant les

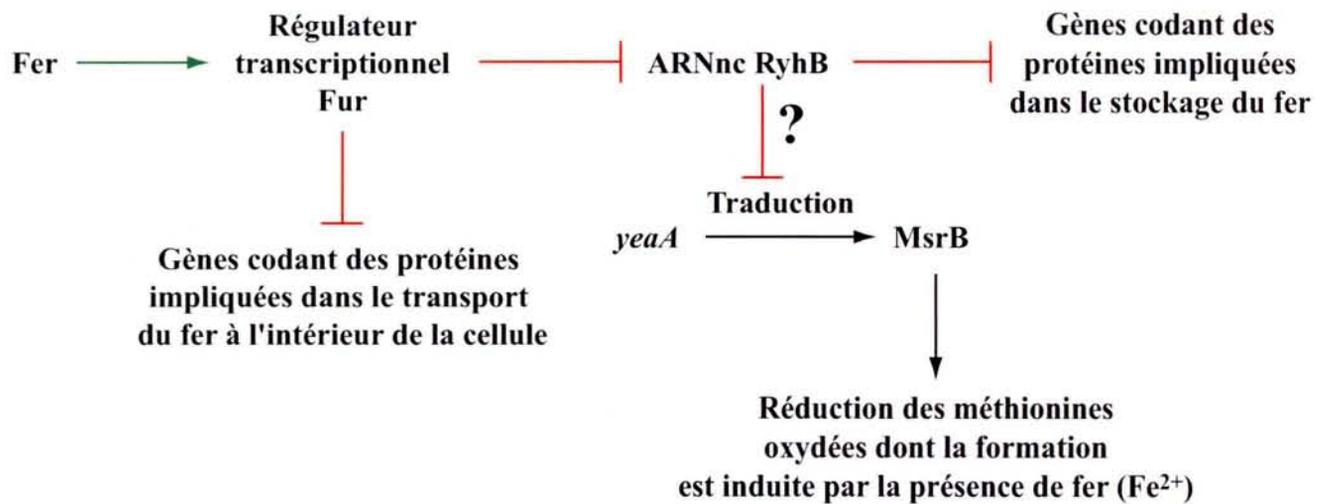


Figure 60 : Modèle de la régulation de l'expression des gènes par le fer chez *E. coli* (d'après Massé et Gottesman, 2002).

Un effet activateur est indiqué par une flèche verte et un effet inhibiteur par un trait rouge. Le rôle inhibiteur potentiel de l'ARN RyhB sur la production de la protéine MsrB est indiqué.

protéines contenant un centre fer-soufre et régule ainsi négativement la synthèse de ces enzymes au niveau traductionnel, ceci en agissant comme un ARN antisens. Actuellement, 6 ARNm cibles de RyhB ont été identifiés, trois d'entre eux codent des enzymes du cycle de Krebs : l'ARNm polycistronique *sdhCDAB*, qui code entre autre la succinate déshydrogénase, l'ARNm *acnA*, qui code l'aconitase A et l'ARNm *fumA*, qui code la fumarase A. Les 3 autres ARNm cibles sont l'ARNm *sodB*, qui code la superoxyde dismutase, l'ARNm *bfr* qui code la bactérioferritine (protéine de stockage du fer) et l'ARNm *finA* qui code la ferritine (protéine de stockage du fer) (Masse et Gottesman, 2002; Geissmann et Touati, 2004). En présence d'ions Fe^{2+} , Fur réprime la transcription de l'ARN RyhB qui ne peut donc plus agir comme un ARN antisens sur ses ARNm cibles. Le taux des enzymes codées par ces ARNm augmente donc dans ces conditions (Masse et Gottesman, 2002).

Le petit ARNnc RyhB agit en association avec la protéine Hfq (11,2 kDa), qui est une protéine très conservée chez les bactéries. Elle participe d'une manière générale à la mise en place des interactions entre beaucoup d'ARNnc et leurs ARNm cibles (Moller *et al.*, 2002) (pour revue : Gottesman, 2004). Cette protéine appartient à une famille de protéines, comprenant les protéines Sm eucaryotes. Les protéines de cette famille forment des multimères (hexamères) associés aux petits ARNnc, ce qui facilite leurs interactions avec leurs cibles (Moller *et al.*, 2002; Moll *et al.*, 2003b; Gottesman, 2004). Comme les protéines Sm, Hfq se fixe préférentiellement sur des séquences ARN en simple brin, riches en résidus A et U et qui sont bordées par des structures tige-boucle (Moller *et al.*, 2002; Moll *et al.*, 2003a). L'ARN RyhB possède un site, ayant ces propriétés, capable de fixer la protéine Hfq. Cette fixation augmente sa stabilité, car elle le protège contre l'action de la RNase E (Moll *et al.*, 2003a; Masse *et al.*, 2003a). Comme la protéine Hfq peut aussi avoir des sites de fixation sur les ARNm cibles, elle peut en interagissant avec les deux partenaires, ARNm et ARNnc, favoriser leur interaction (pour revue : Storz *et al.*, 2004). Ainsi, dans le cas de l'ARNm *sodB*, Hfq facilite l'interaction avec RyhB en se fixant sur l'ARNm. Elle modifie, dans ce cas, la structure de l'ARNm et ainsi facilite la reconnaissance par l'ARNnc RyhB (Geissmann et Touati, 2004). Hfq possède aussi des sites de fixation sur les ARNm et peut aussi avoir un rôle de régulateur post-transcriptionnel sans qu'il y ait forcément implication d'ARNnc. Ainsi, la fixation de Hfq stimule la traduction de l'ARNm *rpoS*, alors qu'elle inhibe la traduction de l'ARNm *ompA*, ceci en bloquant la fixation des ribosomes sur l'ARNm (pour revue : Masse *et al.*, 2003b).

L'équipe de F. Barras avait débuté une étude de l'implication possible de l'ARN régulateur RyhB dans la régulation du gène *yeaA* en présence de fer. Nous avons réalisé des expériences complémentaires, ceci en tenant compte aussi du rôle possible de Hfq.

V.2) Recherche de complémentarités possibles entre l'ARNc RyhB et l'ARNm *yeaA*.

Dans un premier temps, nous avons recherché si l'ARN RyhB présentait des complémentarités de séquence avec la partie 5' de l'ARNm *yeaA* initié au site P1 (Fig. 61). La structure 2D de RyhB a été déterminée en solution grâce à l'utilisation d'une batterie de sondes enzymatiques (Fig. 61 A) (Geissmann et Touati, 2004). Nous avons détecté l'existence de 3 éléments de séquence de l'ARN RyhB complémentaires à la région 5' UTR de l'ARNm *yeaA*. Une première interaction Int 1 impliquerait la structure tige boucle 2 (TB2) de RyhB et les 12 nucléotides en 5' de l'ARNm *yeaA* (positions +1 à +12). Les 12 appariements de bases formés sans interruption ne contiennent qu'une paire G-U, l'énergie libre correspondante calculée d'après le logiciel Mfold à 37°C et dans 1 M NaCl est de -16,3 kcal/mol. Il s'agit d'une interaction potentielle très stable. Sa mise en place implique seulement l'ouverture de quelques paires de bases de la structure TB2 de l'ARN RyhB, sa formation est donc thermodynamiquement favorable. La seconde interaction possible Int 2 impliquerait une région située des positions +26 à +38 par rapport au site d'initiation de la transcription P1 de l'ARNm *yeaA* et toujours la structure TB2 de l'ARN RyhB. Cette seconde interaction potentielle est aussi prédite être stable par Mfold (-11,7 kcal/mol). Elle comporte en effet aussi 12 paires de bases avec un seul résidu en boucle latérale. Elle ne met en œuvre que des résidus en simple brin de la structure TB2 de l'ARN RyhB, ce qui est extrêmement favorable. Il est à noter que la mise en place des deux interactions Int 1 et Int 2 nécessite deux molécules d'ARN RyhB, puisque les séquences de cet ARN mises en jeu sont chevauchantes. A noter aussi que la mise en place d'une interaction telle que Int 2 doit conduire à un blocage complet de la traduction, puisque la séquence RBS et l'AUG initiateur de l'ARNm sont impliqués dans l'interaction. Une troisième interaction possible Int 3 impliquerait une région de *yeaA* située des positions +49 à +57 et une région de l'ARN RyhB qui est structurée en double brin. Cette interaction est de plus faible énergie (ΔG -1,3 kcal/mol) et nécessite l'ouverture d'une structure de l'ARN RyhB. Sa mise en place est donc peu probable. Enfin, une quatrième interaction possible Int 4 impliquerait une région située à proximité de l'extrémité 3' de l'ARNm *yeaA* (positions +347 à +359, Fig. 61) et la séquence située à l'extrémité 5' de

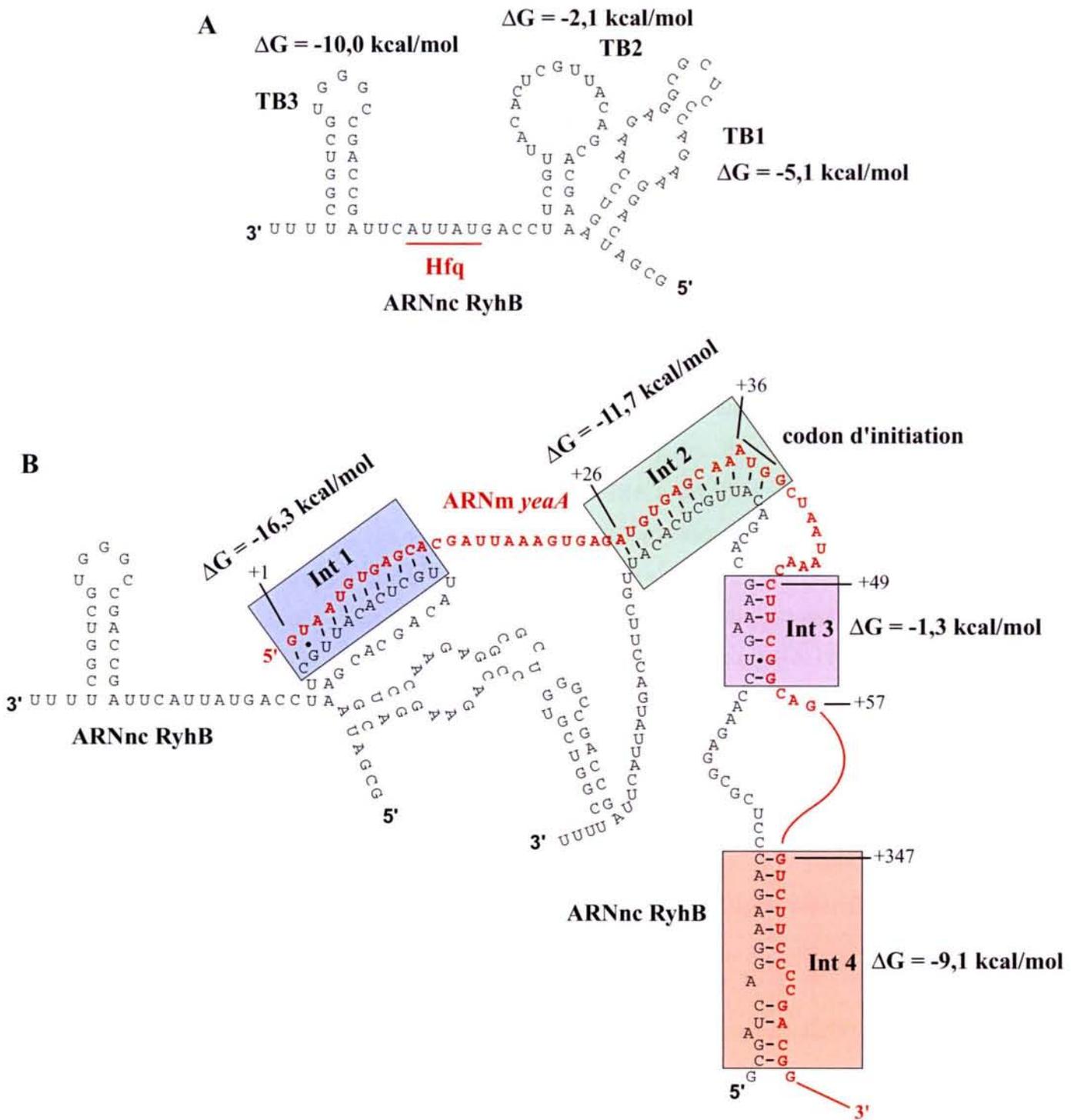


Figure 61 : Interactions possibles entre l'ARNnc RyhB et l'ARNm *yeaA* d'*E.coli*.

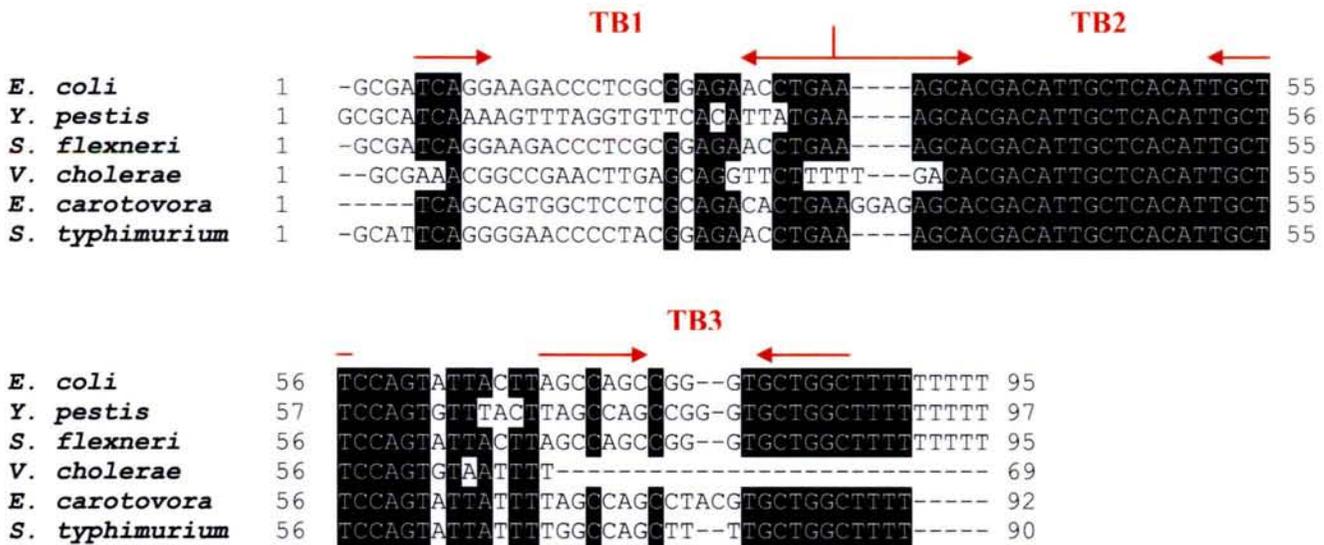
A) Structure 2D de l'ARNnc RyhB (Geissmann et Touati, 2004). Les trois structures tige boucle sont représentées, ainsi que le site potentiel de fixation de la protéine Hfq. L'énergie libre de chaque structure tige boucle à 37°C dans 1 M NaCl a été calculée avec le logiciel MFold (<http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html>).

B) Appariements possibles entre l'ARN RyhB et l'ARNm *yeaA*. L'ARNm *yeaA* est représenté en rouge et les deux molécules RyhB en noir. Les quatre zones potentielles d'interaction (Int 1 à Int 4) sont indiquées par des rectangles de couleurs différentes. Le codon d'initiation de la traduction de l'ARNm *yeaA* est indiqué par un trait noir. L'énergie libre de chaque interaction a été calculée comme pour le panneau A).

l'ARNnc. L'énergie libre de l'interaction possible Int 4 est de -9,1 kcal/mol. Sa mise en place nécessiterait l'ouverture de la structure TB1 de RyhB dont l'énergie libre est de 5,1 kcal/mol. On pourrait imaginer qu'une seule molécule de l'ARNnc RyhB établisse les trois interactions Int 2, Int 3 et Int 4. Jusque là, les études expérimentales sur les ARNnc régulant la traduction, bien que limitées, suggèrent que chez les procaryotes, ces ARNnc s'apparient avec la région 5'UTR des messagers et l'extrémité 5' terminale de l'ORF. Il semble donc que les deux interactions potentielles Int 1 et surtout Int 2, qui bloqueraient la traduction, puissent vraisemblablement être mises en œuvre pour assurer la régulation du gène *yeaA* par l'ARN RyhB. Ceci d'autant plus pour Int 2, qu'elle ne nécessite pas du tout l'ouverture de la structure de l'ARNnc. Il est intéressant de noter que dans le cas de la régulation de l'expression du gène *sodB* par l'ARN RyhB, c'est la même région de la tige boucle 2 de l'ARN RyhB qui est aussi mise en œuvre dans l'interaction et l'appariement a aussi lieu avec la région de l'ARNm contenant la séquence RBS et le codon d'initiation. L'appariement mis en place ne comporte que 9 pb dans ce cas (Geissmann et Touati, 2004). L'interaction entre l'ARN RyhB et l'ARNm *sdhCDAB* semble plus complexe, elle impliquerait la formation d'un grand nombre d'appariements (34 appariements de bases de type Watson Crick non consécutifs) (Masse et Gottesman, 2002), mais dans ce cas, seule la séquence RBS est impliquée dans la formation des appariements. Cette interaction complexe avec l'ARN *sdhCDAB* impliquerait l'ouverture complète de la structure de l'ARN RyhB (Masse et Gottesman, 2002). Il semblerait donc que l'interaction entre l'ARN RyhB et l'ARNm *yeaA* soit plus semblable à celle mise en place dans le cas de l'ARNm *sodB*, que dans le cas de l'ARNm *sdhCDAB*.

Il était intéressant de savoir si la possibilité de former les interactions Int 1 et Int 2 était conservée dans les bactéries. Pour cela, nous avons d'abord recherché l'existence de gènes potentiels d'ARN RyhB dans les génomes séquencés de γ -protéobactéries, ceci par recherche de similarités de séquence avec le logiciel Blast. Nous observons l'existence d'une région d'ADN pouvant coder la structure tige boucle 2 de l'ARN RyhB dans toutes les espèces étudiées et le degré de conservation de séquence est très élevé (Fig. 62 A). Par contre, les séquences environnantes supposées coder les segments formant les structures tige boucle 1 et 3 sont uniquement conservées chez les espèces très proches d'*E. coli* (*Enterobacteriaceae*), à savoir *Shigella flexneri*, *Salmonella typhimurium* et *Yersinia pestis*. Chez les *Vibrionacea* (*Vibrio cholerae*), on note une forte divergence (Fig. 62 A). Ceci renforce l'idée d'un rôle important de la structure tige boucle 2 de l'ARN RyhB. Il est à noter que nous n'avons pas

A



B

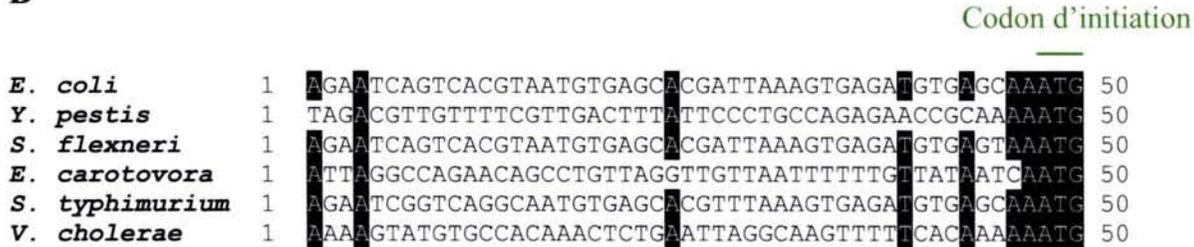


Figure 62 : Alignement des séquences ADN codant potentiellement l'ARN RyhB (A) et de la région située en amont de l'ORF *yeaA* chez différentes espèces de γ -protéobactéries. Les bases identiques entre les différentes espèces sont indiquées par un fond noir. Dans le panneau A, la position des structures tiges boucles identifiées dans l'ARN RyhB d'*E. coli* (Geissmann et Touati, 2004) sont indiquées. Dans le panneau B, le codon d'initiation, indiqué en vert, est aligné à droite.

trouvé d'homologies significatives dans les génomes d'espèces bactériennes n'appartenant pas aux γ -protéobactéries, donc si un ARN équivalent à RyhB existe dans ces espèces, il est très divergent. Par ailleurs, en ce qui concerne les capacités de l'ARN RyhB à former les appariements de bases Int 1 et Int 2 avec l'ARN *yeaA*, le degré de conservation est nettement plus faible, puisque l'alignement des séquences localisées en amont du gène *gapA* chez les γ -protéobactéries, ne met en évidence une forte conservation de la région 5' UTR du gène *yeaA*, que chez les espèces proches d'*E. coli* (Fig. 62 B). Donc l'ARN RyhB ne semble conservé qu'au niveau de la tige boucle 2 et ceci que chez les γ -protéobactéries. Par ailleurs, son interaction avec l'ARNm *yeaA* pourrait être spécifique des *Enterobacteriaceae*. A titre comparatif, nous avons fait l'analyse de la conservation des capacités d'interaction entre l'ARN RyhB et l'ARNm *sodB* et nous faisons la même constatation. Il n'y a conservation des possibilités d'interaction que chez les *Enterobacteriaceae*.

Il est donc possible que ces petits ARN régulateurs soient apparus tard dans l'évolution des bactéries ou aient divergé rapidement au cours de l'évolution.

Sur la base de la forte conservation de séquence de la tige boucle 2 de l'ARNnc RyhB et des forts appariements qu'elle peut former avec la région 5'UTR de l'ARNm *yeaA* d'*E. coli* et en tenant compte de l'observation de l'équipe de F. Barras d'une limitation possible de l'expression de l'ARNm *yeaA* par les ions Fe^{2+} , il nous a paru dans un premier temps intéressant de tester l'effet d'une surexpression de l'ARN RyhB sur l'expression du gène *yeaA* *in vivo*. Cette approche avait été utilisée par Massé *et al.* (2002) pour mettre en évidence le rôle régulateur de RyhB sur l'expression du gène *sdhCDAB*.

V.3) Effet de la surexpression de l'ARN RyhB sur l'expression du gène *yeaA*.

Nous avons surproduit l'ARNnc RyhB dans la souche *E. coli* XL1 blue. Pour cela, nous avons cloné la séquence ADN codant l'ARN RyhB sous contrôle de son propre promoteur dans la cassette de clonage du plasmide pUC19. Le fragment d'ADN à insérer a été amplifié par PCR. Le plasmide recombinant est désigné pUC::*ryhB*. Les cellules XL1 blue transformées par ce plasmide ont été cultivées en milieu LB jusqu'à une DO_{600nm} de 0,5, puis centrifugées et lavées dans le tampon A (EDTA 5 mM, saccharose 25 %, Tris-HCl 50 mM pH 8,0). Les ARN totaux ont été extraits comme décrit dans l'article n°1. Afin d'étudier les variations du taux d'ARNm *yeaA* en fonction du taux d'expression de l'ARN RyhB, nous avons réalisé des

expériences d'extensions d'amorce par la transcriptase inverse avec l'oligonucléotide 437 complémentaire de l'ARNm *yeaA* initié au promoteur P1 (Fig. 63). En témoin, nous avons réalisé une extension avec l'amorce 1043 permettant d'analyser le taux d'ARNm *gapA* initié au promoteur P1 (Fig. 63). L'augmentation du taux d'expression de l'ARN RyhB dans la souche transformée par le plasmide recombinant pUC::*ryhB* a été suivie par la même approche en utilisant l'oligonucléotide 3260 complémentaire de cet ARN (Fig. 63). Nous avons constaté une forte augmentation du taux de l'ARN RyhB (facteur de 50). En parallèle, le taux d'ARNm *yeaA* diminue d'un facteur 1,5 (Fig. 63). La surexpression de l'ARNnc RyhB dans la cellule conduit donc à une légère diminution du taux de transcrits initiés au promoteur *yeaA* P1. L'effet n'est pas très marqué par rapport à la diminution du taux de l'ARNm *sdhCDAB* observé par Massé *et al.* (2002), lorsque l'ARN RyhB est surexprimé.

Ayant observé d'une part une complémentarité entre la région d'initiation de la traduction de l'ARNm *yeaA* et ayant mis en évidence un effet négatif de la surexpression de l'ARN RyhB sur le taux d'ARNm produits par le gène *yeaA*, il nous a paru intéressant de tester expérimentalement les interactions possibles entre l'ARNnc RyhB, l'ARNm *yeaA* et la protéine Hfq *in vitro*. Pour cela, nous avons dû produire la protéine Hfq recombinante et tester son activité.

V.4) Test d'un effet possible de la protéine Hfq sur la formation d'une interaction entre l'ARN RyhB et l'ARNm *yeaA*.

V.4.1) Purification de la protéine recombinante Hfq et vérification de son activité de fixation à l'ARN RyhB.

Nous avons produit et purifié la protéine Hfq en utilisant le protocole utilisé par Vassilieva *et al.* (2003). Pour cela, la souche *E. coli* BL21 DE3 a été transformée par le plasmide pET21b-hfq, qui nous avait été fourni par G. Storz. Après induction par l'IPTG, les cellules ont été centrifugées et les culots cellulaires ont été repris dans le tampon suivant : Tris-HCl 50 mM pH 8,0, 200 mM MgCl₂, 1,5 mM NaCl, 5 mM EDTA. Les cellules ont été lysées par sonication, puis le lysat a été centrifugé 20 min à 14000 g. Comme la protéine Hfq est thermostable, le surnageant a été chauffé à 80°C pendant 20 min pour dénaturer les autres protéines cellulaires. Du fait que ces protéines s'étaient agrégées, elles ont pu être éliminées par centrifugation 40 min à 14000 g. Après ajout de sulfate d'ammonium (1,7 M), le surnageant a été fractionné sur une colonne de chromatographie basée sur la formation

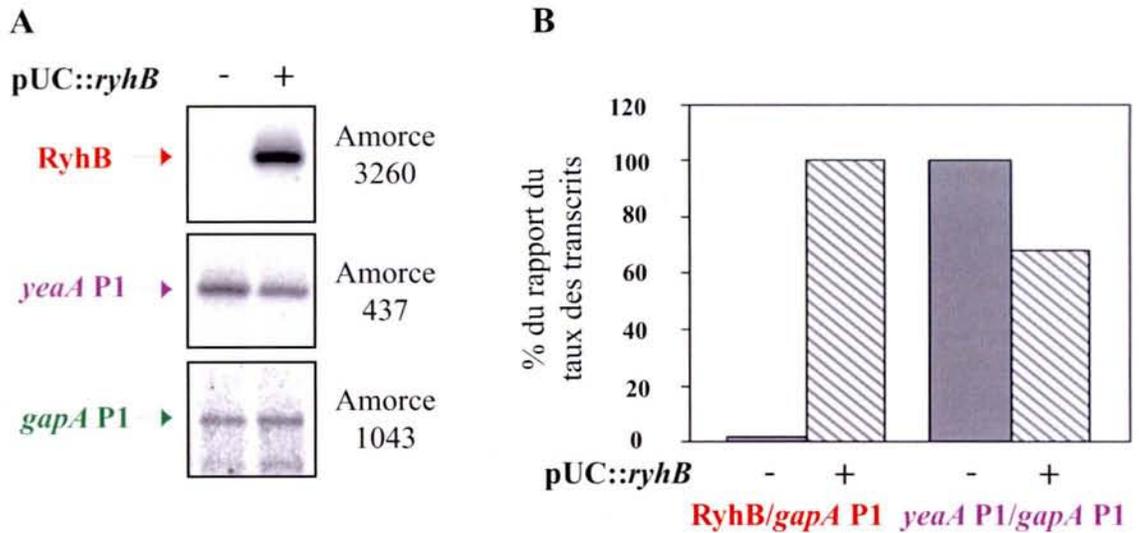


Figure 63 : La surexpression de l'ARNnc RyhB conduit à une diminution du taux d'ARNm *yeaA*.

A) Analyse des taux d'ARN RyhB et d'ARNm *yeaA* et *gapA* P1 par extension d'amorce avec les oligonucléotides 3260, 437 et 1043. Les ARN étaient extraits de cellules d'*E. coli* XL1 blue transformées par le plasmide pUC::*ryhB* (+) ou par le plasmide pUC sans insert (-). Les ADNc correspondants aux différents transcrits sont indiqués.

B) Histogramme représentant l'effet de la présence du plasmide pUC::*ryhB* sur le taux des transcrits RyhB et des transcrits initiés au promoteur *yeaA* P1. La radioactivité présente dans les bandes du gel présenté dans le panneau A a été quantifiée par un PhosphorImager et pour chaque essai le taux d'ARN (RyhB ou *yeaA* P1) a été rapporté à la quantité de transcrits *gapA* P1. Les valeurs obtenues pour les cellules transformées avec le plasmide pUC::*ryhB* ou pUC ont été arbitrairement prises comme référence (100 %).

d'interactions hydrophobes (HIC) HiTrap Butyl FF de 1 ml (Amersham Pharmacia Biotech) (Vassilieva *et al.*, 2003). Après lavage de la colonne avec un tampon contenant 1,7 M de sulfate d'ammonium et 1,5 M de NaCl, les protéines ont été éluées par un gradient linéaire décroissant de sulfate d'ammonium (1,7-0 M) en présence de 1,5 M NaCl et de Tris-HCl 50 mM pH 8,0. Les fractions contenant la protéine Hfq (détection par suivi de la DO_{280nm} et par gel SDS-PAGE) ont ensuite été rassemblées.

Nous avons tout d'abord testé si la protéine Hfq purifiée était bien active. Pour cela, nous avons produit des ARN RyhB uniformément marqués au ³²P, ceci par transcription par l'ARN pol du phage T7, d'une matrice ADN que nous avons amplifiée par PCR en utilisant un oligonucléotide amorce générant un promoteur reconnu par l'ARN pol du phage T7. Ces transcrits RyhB ont été purifiés par électrophorèse, puis utilisés pour réaliser des expériences de retard sur gel avec la protéine Hfq recombinante, en présence d'ARNt à une concentration de 20 ng/μl et dans le tampon suivant (Tris-HCl 10 mM pH 8; MgCl₂ 10 mM; NaCl 50 mM; KCl 50 mM) (Fig. 64 A). Nous avons utilisé des concentrations croissantes en protéine Hfq allant de 0,2 à 3 μM. Les incubations ont été réalisées à 37°C et les fractionnements ont été effectués en gel de polyacrylamide non dénaturant à 6 % à température ambiante.

Lorsque la concentration de la protéine Hfq augmente, la bande correspondant à l'ARN RyhB présente une migration retardée, indiquant la formation du complexe Hfq-RyhB. Le Kd apparent calculé à partir de la mesure de la radioactivité de l'ARN libre et de l'ARN lié (50 % d'ARN sous forme de complexe RyhB-Hfq) est de 1,6 μM. Cette valeur est en accord avec le Kd apparent de 1,5 μM obtenu par Geissmann et Touati (2004) pour cette même interaction. La bande retardée disparaît lorsque l'on réalise des expériences de compétition avec de l'ARN RyhB non marqué. Un rapport molaire de 50 entre l'ARN marqué et l'ARN non marqué conduit à une dissociation complète du complexe Hfq-RyhB (Fig. 64 B). Sur la base de ces données, nous avons conclu que la protéine Hfq que nous avons purifiée était fonctionnelle et que nous pouvions l'utiliser pour l'étude des interactions *yeaA*/RyhB.

V.4.2) La protéine Hfq interagit *in vitro* avec l'ARNm *yeaA*.

Nous avons ensuite utilisé la même technique de retard sur gel pour tester si la protéine Hfq interagit avec l'ARNm *yeaA*. Les transcrits *yeaA* initiés au promoteur P1 ont aussi été produits par transcription *in vitro* en utilisant l'ARN pol T7 et une matrice ADN amplifiée par PCR. Dans un premier temps, nous avons produit un transcrit court allant de la position +1 du promoteur *yeaA* P1 à la position +123. Ce transcrit contenait les séquences Int 1, Int 2 et Int 3

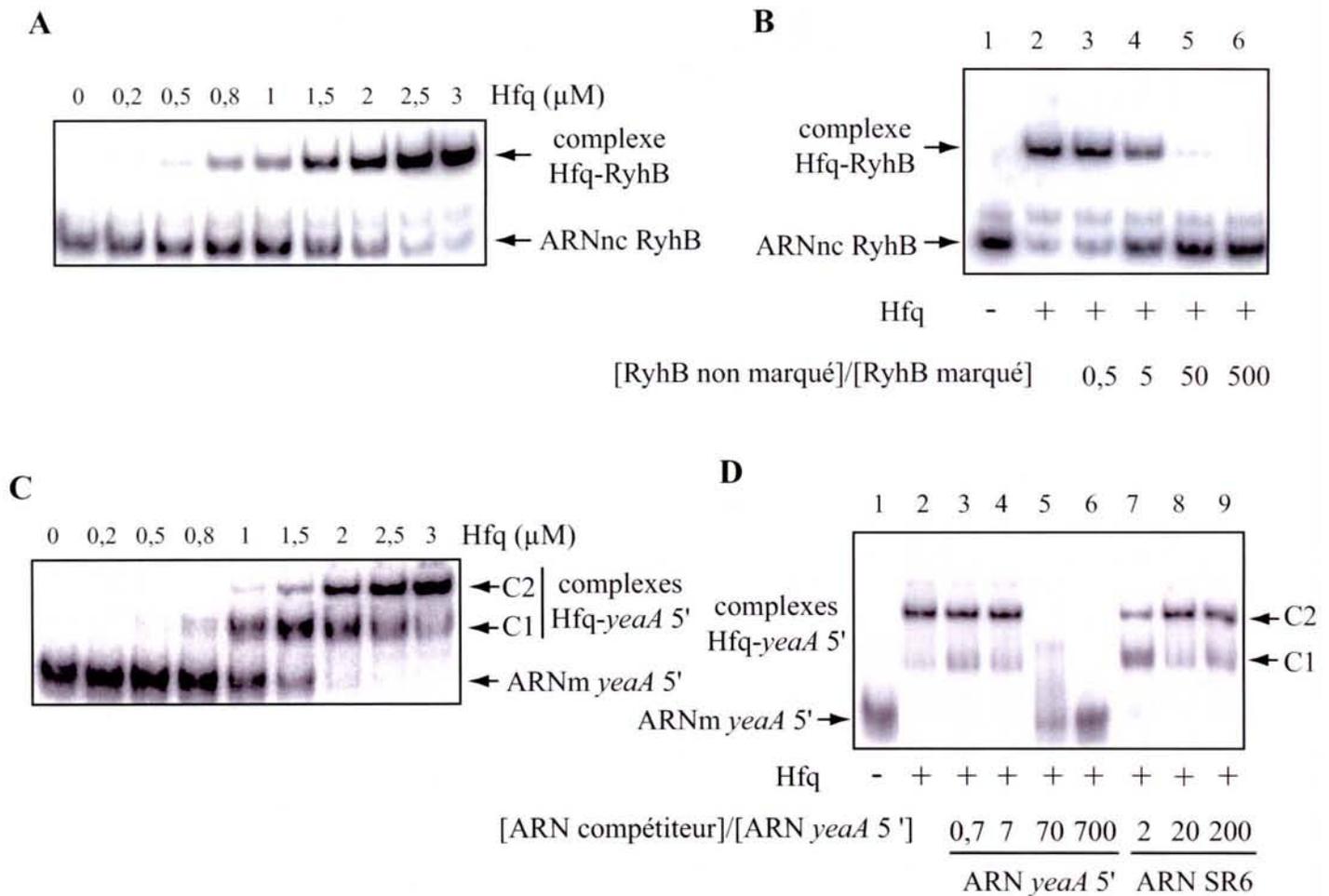


Figure 64 : Mise en évidence d'interactions entre l'ARNm *yeaA* 5' et la protéine Hfq par des expériences de retard sur gel.

A) Etude de l'interaction entre l'ARN RyhB et la protéine Hfq. L'ARNnc RyhB marqué uniformément lors de la transcription a été incubé à une concentration de 10 nM dans le tampon Tris-HCl 10 mM pH 8; MgCl₂ 10 mM; NaCl 50 mM; KCl 50 mM pendant 10 min à 37°C en présence de la protéine Hfq à des concentrations croissantes indiquées au-dessus des pistes. Le transcrit libre et le complexe formé sont indiqués à droite du gel par des flèches.

B) Expériences de compétition réalisées avec l'ARN RyhB froid. Un mélange a été effectué entre l'ARN RyhB uniformément marqué (10 nM) et la protéine Hfq (3 μM) dans les mêmes conditions que dans le panneau A. En même temps, l'ARN RyhB froid a été ajouté à des concentrations croissantes, ceci dans un excès molaire par rapport à l'ARN radioactif allant de 0,5 à 500, comme indiqué sous chaque piste. Les complexes formés ont été fractionnés en gel non dénaturant à 6 %.

C) Etude de l'interaction entre l'ARN *yeaA* 5' uniformément marqué et la protéine Hfq. Dans cette expérience, l'ARN *yeaA* 5' était uniformément marqué lors de la transcription et utilisé à une concentration de 10 nM. Des complexes ont été formés en présence de la protéine Hfq utilisée à des concentrations allant de 0,2 à 3 μM comme indiqué au-dessus des pistes. Les conditions d'incubation étaient les mêmes que dans le panneau A et les complexes ont été fractionnés en gel non dénaturant à 6 %. Deux complexes C1 et C2 ont été détectés dans cette expérience.

D) Expérience de compétition réalisée avec l'ARN *yeaA* 5' froid ou l'ARN SR6 de *P. abyssi*. La protéine Hfq a été utilisée à une concentration de 3 μM . L'ARN marqué *yeaA* 5' était toujours utilisé à une concentration de 10 nM. L'ARN *yeaA* 5' froid, ou l'ARN SR6 froid ont été ajoutés dans des rapports [ARN compétiteur]/[ARN *yeaA* 5'] allant de 0,7 à 200 comme indiqué sous les pistes.

susceptibles d'interagir avec l'ARNnc RyhB. Cet ARN de 123 nt (10 nM) appelé *yeaA* 5' a été incubé 10 min à 37°C en présence de protéine Hfq, ceci avec des concentrations croissantes en protéine allant de 0,2 à 3 µM. Le fractionnement en gel de polyacrylamide non dénaturant des produits formés a permis de mettre en évidence la formation d'au moins deux complexes ARN-protéine (C1 et C2) distincts (Fig. 64 C). Le Kd apparent du complexe C1 est d'environ 0,9 µM. A partir d'une concentration de protéine Hfq de 1,5 µM, ce complexe disparaît au profit du complexe C2 dont le Kd apparent peut être estimé à 2 µM. Il pourrait donc exister plusieurs sites de fixation de la protéine Hfq sur l'ARN correspondant à l'extrémité 5' de l'ARNm *yeaA*, l'un ayant une plus forte affinité que l'autre. A faible concentration en protéine Hfq, seul le site de plus forte affinité serait occupé (complexe C1), alors qu'à forte concentration en protéine Hfq, les deux sites seraient occupés (complexe C2). Néanmoins, une autre possibilité que nous ne pouvons par exclure est que le complexe C2 corresponde à la dimérisation de la protéine Hfq, chaque molécule de protéine se liant à une molécule différente d'ARN *yeaA* 5' puis dimérise, ou tout simplement qu'à une concentration élevée en protéine Hfq, une deuxième protéine s'associe à celle déjà fixée sur l'ARN *yeaA* 5'. Les Kd apparents obtenus pour la formation des deux complexes C1 et C2 sont très supérieurs aux Kd apparents obtenus pour la fixation de la protéine Hfq sur l'ARNm *sodB* (1,8 nM) (Geissmann et Touati, 2004), ce qui suggère que l'interaction entre Hfq et la région 5' de l'ARNm *yeaA*, si elle a lieu *in vivo*, ne doit avoir lieu qu'en présence d'une forte expression de la protéine Hfq.

Nous avons réalisé les mêmes expériences de compétition avec de l'ARN froid que celle réalisée pour l'ARN RyhB. Pour cela, l'ARN *yeaA* 5' a été produit sans incorporation de nucléotide radioactif. Nous avons ensuite ajouté des concentrations croissantes de ce transcrit non marqué (de 7 nM à 7 µM) au mélange réactionnel de formation du complexe Hfq-*yeaA* (concentration en protéine Hfq de 3 µM, concentration en ARN *yeaA* 5' radioactif de 10 nM, incubation à 37°C) (Fig. 64 D). Le complexe *yeaA*-Hfq était complètement dissocié lorsque le rapport entre ARN marqué et ARN froid était de 70 (Fig. 64 D, pistes 3 à 6). Pour tester la spécificité de l'interaction, nous avons utilisé en tant qu'ARN compétiteur, un ARNnc de type sARN à boîte C/D de *Pyrococcus abyssi*, appelé SR6. Dans ce cas, le complexe *yeaA*-Hfq n'était pas dissocié, même pour le rapport molaire le plus élevé d'ARN SR6/ARN *yeaA* 5' (200) (Fig. 64 D, pistes 7 à 9), ce qui suggère que l'interaction détectée entre Hfq et l'ARN *yeaA* est spécifique.

V.4.3) La protéine Hfq ne semble pas faciliter l'interaction entre la région 5' de l'ARNm *yeaA* et l'ARNnc RyhB.

Dans un premier temps, nous avons testé par des expériences de retard sur gel en incubant ensemble les trois partenaires (ARN RyhB, ARNm *yeaA* et la protéine Hfq), 10 min à 37°C dans le même tampon que précédemment, si des complexes étaient formés. Un test témoin a été réalisé en parallèle avec seulement l'ARN RyhB marqué (10 nM) et la protéine Hfq (3 µM). L'interaction des trois partenaires supposés a été testée par incubation avec de l'ARN RyhB marqué (10 nM), la protéine Hfq (3 µM) et l'ARN *yeaA* 5' non marqué utilisé à des concentrations croissantes (rapports molaires ARN *yeaA* 5'/ARN RyhB de 7 à 70) (Fig. 65 A). A la concentration la plus élevée, un complexe plus fortement retardé que celui formé avec seulement l'ARN RyhB et la protéine Hfq est formé (piste 4, Fig. 65 A). Nous avons ensuite réalisé l'expérience symétrique, à savoir que l'ARN *yeaA* 5' était marqué (10 nM), et incubé dans les conditions décrites ci-dessus, d'une part avec la protéine Hfq seule (3 µM), d'autre part avec cette même protéine (3 µM) et l'ARN RyhB non marqué (50 nM) (rapport molaire de 5) (pistes 1 à 3, Fig. 65 B). Lorsque l'ARN RyhB et la protéine Hfq sont ajoutés, un complexe ayant un retard de migration plus important que celui observé lorsque la protéine Hfq est ajoutée seule, était visible (piste 3).

La question était de savoir si la formation de ce complexe nécessitait la présence de la protéine Hfq. Pour répondre à cette question, nous avons testé si incubés ensemble, l'ARN *yeaA* 5' et l'ARN RyhB étaient capables de former un complexe. L'ARN *yeaA* 5' marqué (10 nM) a été incubé toujours dans les mêmes conditions que ci-dessus au niveau temps, température et tampon, en présence de concentrations croissantes d'ARN RyhB (5 nM, 50 nM et 500 nM) (rapport molaire ARN *yeaA* 5'/ARN RyhB de 0,5, 5 et 50) (pistes 7 à 9, Fig. 65 B). Dans ces conditions, d'une part nous observons un étalement de la bande contenant l'ARN *yeaA* 5' vers des mobilités plus faibles. D'autre part, une bande bien distincte supplémentaire apparaît. L'intensité de cette bande supplémentaire augmente lorsque la concentration en ARN RyhB est augmentée. Différents types de complexes semblent donc se former lorsque l'on incube l'ARN *yeaA* 5' avec l'ARN RyhB. La bande supplémentaire qui augmente avec l'augmentation de concentration d'ARN RyhB contient sans doute un duplex ARN-ARN *yeaA* 5'-ARN RyhB d'après la mobilité électrophorétique observée (piste 9). Afin de savoir si le même type de complexe se formait entre les deux ARN en présence ou en absence de la protéine Hfq, nous avons incubé les deux ARN (ARN *yeaA* 5' marqué (10 nM) et ARN RyhB froid (50 nM, rapport molaire ARN *yeaA* 5'/ARN RyhB de 5) en présence de

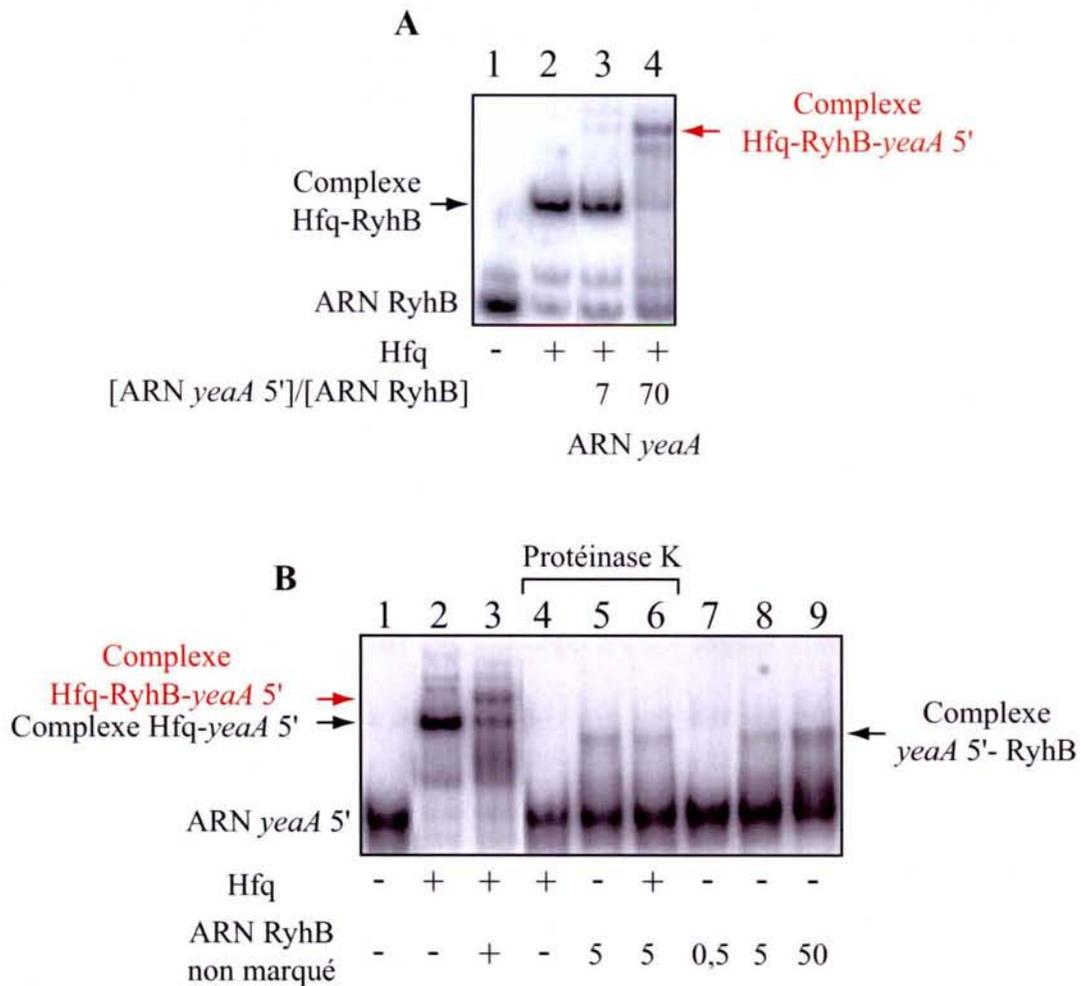


Figure 65 : Etude par expériences de retard sur gel des interactions mises en place *in vitro* entre l'ARN *yeaA* 5', l'ARNnc RyhB et la protéine Hfq.

A) Le complexe formé entre la protéine Hfq et l'ARN est déplacé vers la formation d'un complexe fortement retardé en présence de l'ARN *yeaA* 5'. Les différents échantillons de l'ARN RyhB marqué uniformément lors de la transcription, seul (10 nM) (piste 1), en présence de la protéine Hfq (3 μM) (piste 2), en présence de cette protéine et d'ARNm *yeaA* 5' en excès molaire de 7 x (piste 3) et 70 x (piste 4) par rapport à l'ARN RyhB, ont été incubés 10 min à 37°C dans le tampon Tris-HCl 10 mM pH 8; MgCl₂ 10 mM; NaCl 50 mM; KCl 50 mM et les complexes formés ont été fractionnés en gel de polyacrylamide non dénaturant à 6 %. Les complexes formés avec la protéine Hfq seule et avec cette protéine et l'ARN RyhB sont indiqués par des flèches.

B) Etude des interactions formées par les ARN *yeaA* 5' et RyhB en présence ou en absence de la protéine Hfq. Dans ces expériences, l'ARN *yeaA* 5' était marqué en 5' et utilisé à une concentration de 10 nM. La protéine Hfq était utilisée à une concentration de 3 μM (indiqué par +). L'ARN RyhB était utilisé dans un excès molaire par rapport au transcrit marqué allant de 0,5 à 50 comme indiqué sous les pistes 5 à 9. Pour les pistes 4 à 6, les échantillons ont été traités à la protéinase K et au phénol pour dégrader et éliminer la protéine Hfq. Après ce traitement, les échantillons ont été précipités à l'acétate de sodium et à l'éthanol, puis repris dans le bleu de dépôt.

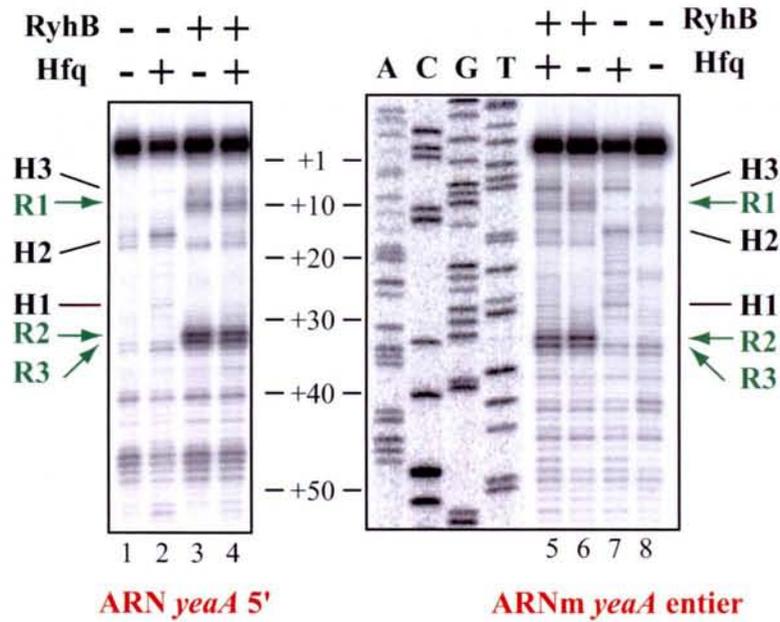
protéine Hfq à une concentration de 3 μ M, puis nous avons traité les complexes formés par la protéinase K (15 min à 37°C). Ceci afin de dégrader la protéine Hfq par cette protéase sans modifier les complexes ARN-ARN formés (pistes 4 à 6, Fig. 65 B). Nous constatons après fractionnement sur gel qu'un même taux de duplex d'ARN est obtenu avec ou sans protéine Hfq, ce qui suggère que cette protéine ne favorise pas la mise en place d'une interaction entre les ARN *yeaA* 5' et RyhB *in vitro*. Ceci pourrait peut-être expliquer l'effet négatif faible observé sur l'expression du gène *yeaA* lors de la surexpression de l'ARN RyhB *in vivo*. Dans ce système, contrairement aux systèmes *sdhCDAB* et *sodB* et aux autres systèmes d'interaction entre ARNnc et ARNm déjà décrits, il n'y aurait pas d'effet conjugué de l'ARN RyhB et de la protéine Hfq.

V.4.4) Localisation des sites d'interactions mises en place *in vitro* entre l'ARNnc RyhB, l'ARN *yeaA* 5' et la protéine Hfq.

Etant donné que nous avons observé une interaction possible entre les ARN *yeaA* 5' et RyhB *in vitro*, il était intéressant de savoir si ces interactions correspondaient aux complémentarités mises en évidence par comparaison des séquences. Par ailleurs, il était aussi intéressant de rechercher les sites possibles de fixation de la protéine Hfq sur la région 5' de l'ARNm *yeaA*, puisque nous avons vu qu'elle se liait à l'ARNm *yeaA*. Pour cela, nous avons utilisé deux approches et nous nous sommes focalisés sur la région 5' UTR de l'ARNm *yeaA*. La première approche était basée sur la technique d'extension d'amorce. En effet, une expérience d'élongation d'amorce réalisée en conditions non dénaturantes (10 min à 37°C) avec l'oligonucléotide 437 complémentaire de l'ARNm *yeaA* sur un transcrit correspondant à tout l'ARNm *yeaA* et sur l'ARN *yeaA* 5', nous avait permis d'observer que la transcriptase inverse fait peu de pauses dans la région 5' de l'ARNm *yeaA* (Fig. 66). Ceci suggérait un faible degré de structuration interne favorable à un appariement avec l'ARN RyhB et à l'interaction avec la protéine Hfq. Par ailleurs, cette situation nous permettait d'essayer d'obtenir des informations sur les capacités d'interaction de cette région 5' de l'ARNm *yeaA* avec l'ARN RyhB et la protéine Hfq par extension d'amorce. On pouvait en effet s'attendre à l'apparition de pauses dans le cas de la mise en place d'une interaction stable avec l'ARN RyhB ou la protéine Hfq.

Les expériences réalisées avec la protéine Hfq seule ne nous ont pas permis de tirer beaucoup d'informations aussi nous les décrivons rapidement. Ces expériences d'extension d'amorce ont été réalisées sur des complexes formés dans les conditions décrites ci-dessus avec

A



B

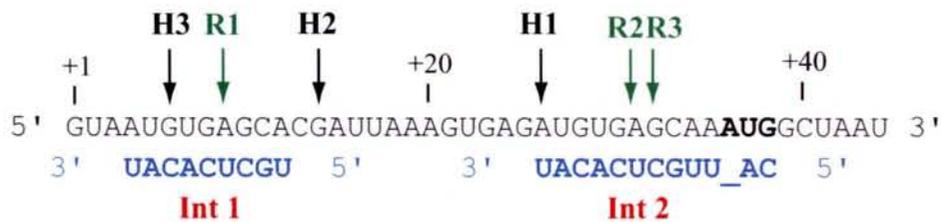


Figure 66 : Recherche d'appariements de bases entre l'ARN *yeaA* 5' et l'ARN RyhB et entre l'ARNm *yeaA* entier et l'ARN RyhB par élongation d'amorce par la transcriptase inverse.

A) L'oligonucléotide amorce 437 a été hybridé sur l'ARN *yeaA* 5' ou l'ARNm *yeaA* entier auquel il est complémentaire. Les hybrides formés ont ensuite été utilisés à une concentration de 70 nM pour l'ARN *yeaA* 5' et 20 nM pour l'ARNm *yeaA* entier, pour former des complexes soit avec la protéine Hfq seule (3 μM), soit avec l'ARN RyhB seul (1 μM ou 3,5 μM), soit avec les deux, ceci dans les mêmes conditions que dans la figure 63. Une extension de l'amorce fixée à l'ARN *yeaA* 5' a ensuite été réalisée par incubation 10 min à 37°C en présence de reverse transcriptase et de dNTP. Les produits d'extension ont été fractionnés sur un gel de séquençage, en parallèle avec les produits d'un séquençage ADN réalisé avec le même oligonucléotide. Les pistes 1 à 4 correspondent à l'analyse de la région 5' du transcrit *yeaA* et les pistes 5 à 8 correspondent à l'analyse du transcrit *yeaA* entier. Les arrêts d'élongation qui apparaissent en présence de l'ARN RyhB sont indiqués par des flèches vertes numérotées R1 à R3 et ceux apparaissant en présence de la protéine Hfq par des traits noirs numérotés H1 à H3.

B) Représentation des résultats obtenus dans le panneau A sur les interactions Int 1 et Int 2 qui peuvent être formées entre les deux ARN. En haut, la séquence en 5' de l'ARNm *yeaA* est indiquée en noir. En-dessous, les séquences complémentaires de l'ARN RyhB sont indiquées en bleu. Les flèches indiquent les positions des arrêts d'élongation de la transcriptase inverse obtenus avec l'ARN RyhB (R1 à R3) ou avec la protéine Hfq (H1 à H3). Le codon d'initiation de la traduction est indiqué en gras.

l'ARNm *yeaA* entier (20 nM) ou l'ARN *yeaA* 5' (70 nM) et la protéine Hfq à la concentration de 3 μ M. Les pauses observées en présence de protéine Hfq n'étaient pas identiques lorsque l'expérience était réalisée sur l'ARN *yeaA* entier ou sur l'ARNm *yeaA* 5' (Fig. 66 A). Trois arrêts étaient observés pour l'ARNm entier, seul l'un de ces 3 arrêts était clairement visible dans le cas de l'ARN *yeaA* 5' (Fig. 66 A). Les pauses détectées dans ce type d'expérience peuvent soit correspondre à un encombrement stérique créé par la protéine fixée, soit à un changement de conformation de l'ARN. Une expérience complémentaire basée sur la technique d'empreinte par les RNases devra être réalisée pour savoir si les pauses observées correspondent à des sites de fixation.

Par contre, l'approche extension d'amorce a donné des résultats beaucoup plus clairs lorsque l'expérience a été réalisée sur les complexes formés entre les ARNm *yeaA* entier ou l'ARN *yeaA* 5' et l'ARN RyhB (Fig. 66 A). Les concentrations d'ARN utilisées pour former les complexes étaient respectivement de 20 nM pour l'ARNm *yeaA* avec l'ARN RyhB à une concentration de 1 μ M et 70 nM pour l'ARN *yeaA* 5' avec l'ARN RyhB à 3,5 μ M, soit dans les deux cas un excès de 50 fois de l'ARN RyhB. Les conditions d'incubation étaient celles décrites ci-dessus. L'oligonucléotide était hybridé sur l'ARN analysé (*yeaA*) avant la formation du complexe. Nous avons observé de fortes pauses de la transcriptase inverse localisées à la bordure des éléments de séquence de la région 5' de l'ARNm *yeaA* qui peuvent s'apparier avec l'ARN RyhB (Fig. 66 B). Ceci était aussi bien observé dans le cas de la région 5' de l'ARN *yeaA*, que dans le cas de l'ARNm entier. De plus, les mêmes pauses de forte intensité étaient observées lorsque les duplex d'ARN étaient formés en présence de la protéine Hfq (Fig. 66 A).

Ces résultats suggéraient fortement que les interactions Int 1 et Int 2 étaient mises en place *in vitro*, ceci aussi bien pour l'ARNm tronqué que pour l'ARNm entier. Pour confirmer ces données, nous avons utilisé la technique d'étude des structures des ARN par hydrolyse ménagée par la RNase T1 (Fig. 67). Cette RNase coupe spécifiquement les liaisons phosphodiester en 3' des résidus G et lorsqu'elle est utilisée en conditions limitantes, elle n'agit que sur les régions en simple brin et donne donc des informations sur la structure 2D des ARN. Pour identifier les positions des coupures, on peut soit marquer l'ARN à une de ses extrémités, soit utiliser l'approche extension d'amorce. Etant donné que les interactions Int 1 et Int 2 sont mises en place à proximité de l'extrémité 5' de l'ARN *yeaA* 5', nous avons choisi de marquer cet ARN en 5' par action de la T4 polynucléotide kinase comme décrit dans la partie Matériels et Méthodes.

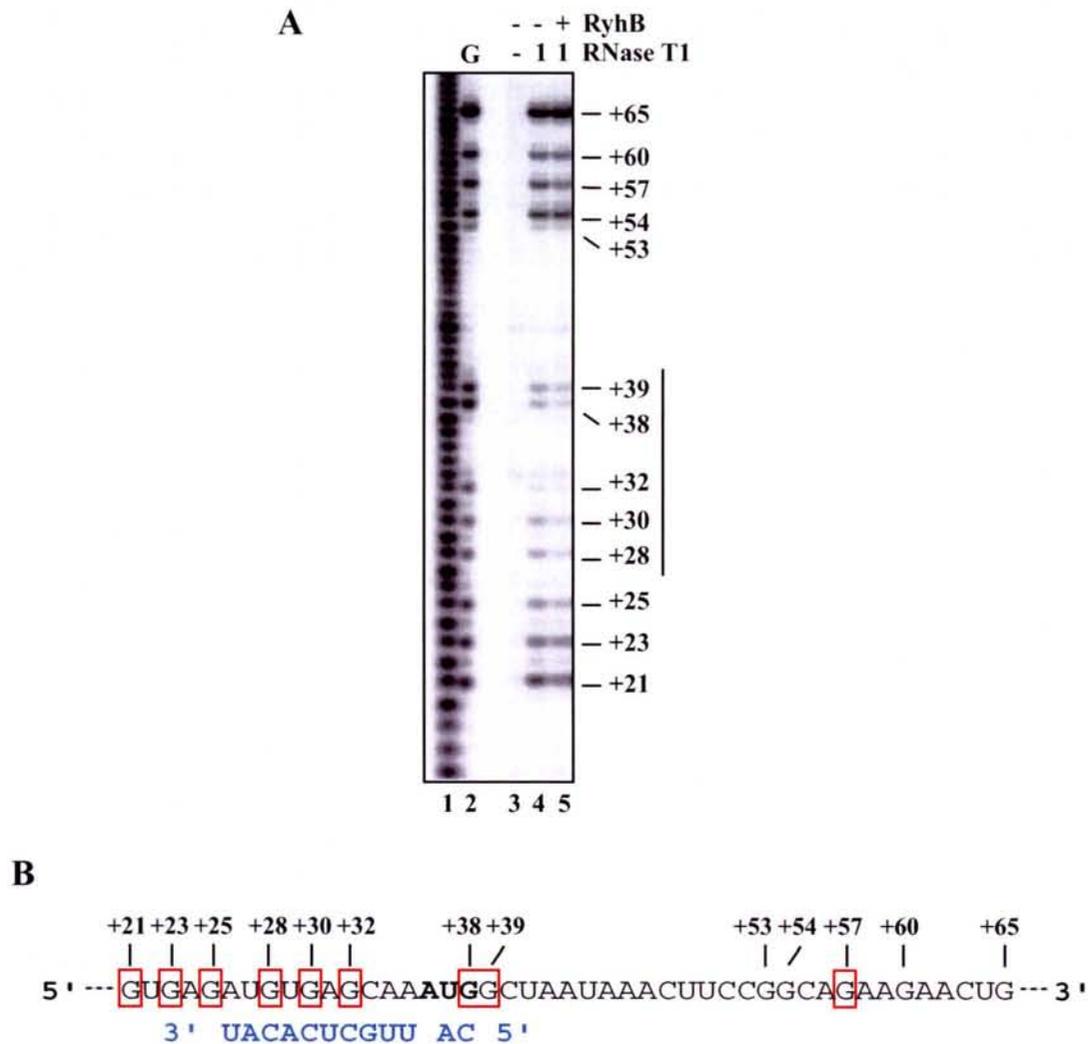


Figure 67 : Etude de l'interaction entre les ARN *yeaA* 5' et RyhB par empreinte à la RNase T1.

A) L'ARN *yeaA* 5' a été marqué en 5', puis incubé avec de la RNase T1 en conditions non dénaturantes (voir les conditions utilisées dans le texte). Les produits d'hydrolyse ont été fractionnés en gel de polyacrylamide dénaturant. La première piste correspond à une dégradation statistique de l'ARN en présence de bicarbonate de soude (pH 9,2) et la deuxième, à une hydrolyse avec une concentration plus forte de RNase T1 et en conditions dénaturantes (urée 8 M), ce qui permet de repérer la position de tous les résidus G.

Dans la piste 5, l'ARN RyhB a été ajouté dans un excès molaire de 50x par rapport à l'ARN *yeaA* 5'. La position des résidus clivés par la RNase T1 par rapport à l'extrémité 5' de l'ARNm est indiquée à droite du gel.

B) Représentation des résultats obtenus par empreinte à la RNase T1 sur la séquence en 5' de l'ARN *yeaA* 5'. Les résidus protégés sont indiqués par un rectangle rouge. La séquence de l'ARN RyhB qui pourrait interagir avec l'ARNm *yeaA* est indiquée en bleu.

Nous avons dans un premier temps réalisé une hydrolyse limitée de l'ARN *yeaA* 5' marqué en 5' par la RNase T1 en conditions non dénaturantes (tampon Tris-HCl 10 mM pH 8; MgCl₂ 10 mM; NaCl 50 mM; KCl 50 mM, 10 min à 20°C, une unité d'enzyme pour 0,15 nM d'ARN). Les produits d'hydrolyse ont été fractionnés par électrophorèse en gel de polyacrylamide dénaturant à 7 %, ceci en parallèle avec les produits d'une hydrolyse de ce même ARN réalisée en conditions dénaturantes (urée 8M) et les produits d'une hydrolyse alcaline générant statistiquement des clivages au niveau de chaque liaison phosphodiester (Fig. 67 A). Comme on peut le constater, presque tous les résidus G compris entre les positions +21 et +65 par rapport au site d'initiation de la transcription étaient clivés dans les conditions non dénaturantes. Seul le résidu +32 était très faiblement clivé. Les clivages aux positions +28, +30, +38 et +39, étaient d'intensité modérée, en comparaison des clivages aux positions +21, +54, +57, +60 et surtout +65. L'ensemble de ces données suggèrent un degré de structuration limité de la région +20 à +65 de l'ARN *yeaA* 5', avec néanmoins l'existence possible d'une structuration plus importante du segment +28 à +39, c'est-à-dire de la séquence complémentaire de l'ARN RyhB dans l'interaction Int 2. L'identification de la structure à laquelle ce segment de l'ARN *yeaA* 5' participe nécessitera l'emploi de l'ensemble de la panoplie de RNases et d'agents chimiques utilisé au laboratoire pour l'étude de la structure 2D des ARN.

Cette étude nous avait néanmoins permis de constater que les résidus G+28, +30, +38 et +39 sont sensibles à l'action de la RNase T1 dans l'ARN *yeaA* 5' en conditions non dénaturantes, nous pouvions donc utiliser ces clivages comme sonde pour tester la mise en place d'une interaction avec l'ARN RyhB. Pour cela, un complexe entre l'ARN *yeaA* 5' marqué en 5' et l'ARN RyhB froid a été formé dans les conditions décrites ci-dessus (rapport molaire de 50) et ce complexe a été hydrolysé par la RNase T1 dans les mêmes conditions que l'ARN *yeaA* 5' seul. Les produits d'hydrolyse de l'ARN seul et du complexe ont été fractionnés en parallèle sur un même gel de séquençage (Fig. 67 A). Une mesure de la radioactivité dans les bandes de gel montre une absence de variation du taux de clivage aux résidus +54, +60 et +65 au sein du complexe. Dans ce segment, on note seulement une légère diminution du taux de clivage à la position +57. Par contre, les taux de clivage de tous les résidus du segment +39 à +21 sont diminués et les diminutions les plus marquées se situent aux positions +28, +30 et +38 (trois résidus qui participent à l'interaction Int 2) (Fig. 67 B). Ces données d'empreinte renforçaient donc l'idée de la mise en place de l'interaction Int 2 *in vitro*. Pour compléter ces données, il sera important de réaliser la même expérience en faisant varier les concentrations relatives des deux ARN. Nous pourrions aussi élargir le panel de sondes de structure utilisées

et réaliser la même analyse sur l'ARN RyhB. Enfin, il sera important de générer des mutations dans les deux séquences partenaires de l'interaction Int 2, ceci pour tester les effets produits sur les complexes formés entre les deux ARN.

V.5) Perspectives pour la poursuite de cette étude.

Bien que cette étude sur la régulation de la traduction de l'ARNm *yeaA*, que j'ai entreprise à la fin de mes travaux de thèse, ne soit pas terminée, les résultats que j'ai obtenus apportent des informations intéressantes. Ils pourront être les bases d'une étude plus complète qui pourra être poursuivie après ma thèse au laboratoire.

En effet, nous avons observé l'existence d'une forte complémentarité entre deux segments de la région 5' UTR de l'ARN *yeaA* P1 et la boucle terminale de la tige boucle 2 de l'ARN RyhB. Nos données expérimentales suggèrent fortement que les interactions correspondantes sont mises en place *in vitro*. Enfin, la surexpression de l'ARN RyhB s'accompagne d'une diminution du taux de l'ARNm *yeaA* d'un facteur 1,5. Ceci pourrait être lié à la dégradation d'une partie de cet ARN dont la traduction serait bloquée par fixation de l'ARN RyhB au site Int 2. *In vitro*, la protéine Hfq ne favorise pas l'interaction entre l'ARNm *yeaA* et l'ARN RyhB, vraisemblablement parce que, comme nous l'avons constaté, cette protéine se fixe à l'ARNm *yeaA* mais avec une assez faible affinité.

On pourrait donc envisager différents modes de régulation de la traduction des ARNm par l'ARN RyhB, avec des situations où, à la fois l'ARN RyhB et la protéine Hfq jouent un rôle important dans la formation d'un duplex d'ARN bloquant la traduction. C'est le cas pour les ARNm *sdhCDAB* et *sodB* (Masse et Gottesman, 2002; Geissmann et Touati, 2004). Dans d'autres cas, le duplex d'ARN serait mis en place sans intervention marquée de la protéine Hfq. Il est possible d'ailleurs que dans ce cas, une ou plusieurs protéines autres que Hfq soi(en)t impliquées dans la mise en place de l'interaction. La surexpression de l'ARN RyhB sans surexpression de ces protéines pourrait alors ne conduire qu'à une faible réduction du taux de l'ARNm cible. Il s'agit bien sûr de pistes pour de futures recherches sur le système ARNm *yeaA*/ARN RyhB.

L'hypothèse d'un blocage de la traduction devra être testée par des essais de traduction *in vitro* dans un extrait S30 d'*E. coli*, ceci en testant si l'ajout de l'ARN RyhB dans le milieu réactionnel bloque la traduction d'un ARNm *yeaA* produit *in vitro*. Si la réponse est positive, il sera intéressant de compléter l'étude par emploi de l'approche de « toeprint ». Cette technique consiste à tester si la formation d'un complexe ARNm-sous-unité 30S du ribosome

(avec ou sans l'ARNt initiateur) est bloquée par ajout de l'ARN régulateur. Le test est réalisé par élongation d'amorce par la transcriptase inverse, en effet l'enzyme s'arrête au niveau du complexe fixé sur la séquence RBS et le codon initiateur. Si l'ARN RyhB est bien un inhibiteur de la traduction, on s'attend à ce que l'intensité de la pause observée au niveau du complexe diminue. Il sera aussi intéressant d'utiliser des extraits S30 préparés à partir de souches d'*E. coli* cultivées en présence ou en absence de fer et de réaliser des tests de traduction *in vitro* d'ARNm *yeaA* produits par transcription *in vitro*.

Enfin, pour démontrer qu'il existe bien une interaction entre l'ARN RyhB et l'ARNm *yeaA* *in vivo*, nous envisageons l'utilisation d'une technique développée au laboratoire pour l'étude de l'épissage de l'ARN du virus HIV-1, ceci en collaboration avec l'équipe du Pr Luhrmann du Max Planck Institut de Gottingen. Cette technique consiste à produire l'ARN étudié sous forme d'un ARN hybride contenant un site de fixation pour un antibiotique, la tobramycine (Fig. 68) (Cho *et al.*, 1998; Hartmuth *et al.*, 2002). Cet ARN hybride associé à ses partenaires cellulaires peut alors être purifié sur une colonne de chromatographie sur laquelle a été greffée la tobramycine. Nous avons déjà réalisé une construction permettant d'exprimer chez *E. coli*, l'ARN RyhB fusionné à l'aptamère reconnaissant la tobramycine. Nous avons vérifié que l'ARN fusionné exprimé dans les cellules est bien capable de se fixer sur des billes de sépharose sur lesquelles nous avons greffé la tobramycine. Cette technique pourra nous permettre de mettre en évidence les partenaires associés à l'ARN RyhB, qu'ils soient de nature ARN ou de nature protéique.

Nos données suggèrent que l'ARN RyhB, dont l'expression est diminuée en présence de fer (Fe^{2+}) dans le milieu de culture du fait de la répression de sa transcription par la protéine Fur, inhibe la traduction de l'ARNm *yeaA*. Quelle signification biologique pourrait avoir ce résultat ?

Les tests effectués au laboratoire par l'équipe Enzymologie Moléculaire suggèrent qu'un atome de fer est présent dans la protéine MsrB (A. Olry et G. Branlant, non publié). Le site de fixation du métal dans la protéine pourrait accepter un atome de Zinc (Zn^{2+}) en absence de complémentation du milieu par du fer (FeSO_4). Dans ces conditions, 80% des protéines seraient liées à un ion Zn^{2+} . Par contre, en présence de concentrations croissantes en fer (FeSO_4), jusqu'à 40% des protéines MsrB seraient liées au fer (A. Olry et G. Branlant, non publié). Ces résultats suggèrent que la protéine MsrB pourrait servir au stockage du fer lorsque celui-ci est en forte concentration dans la cellule et donc la régulation de l'ARNm

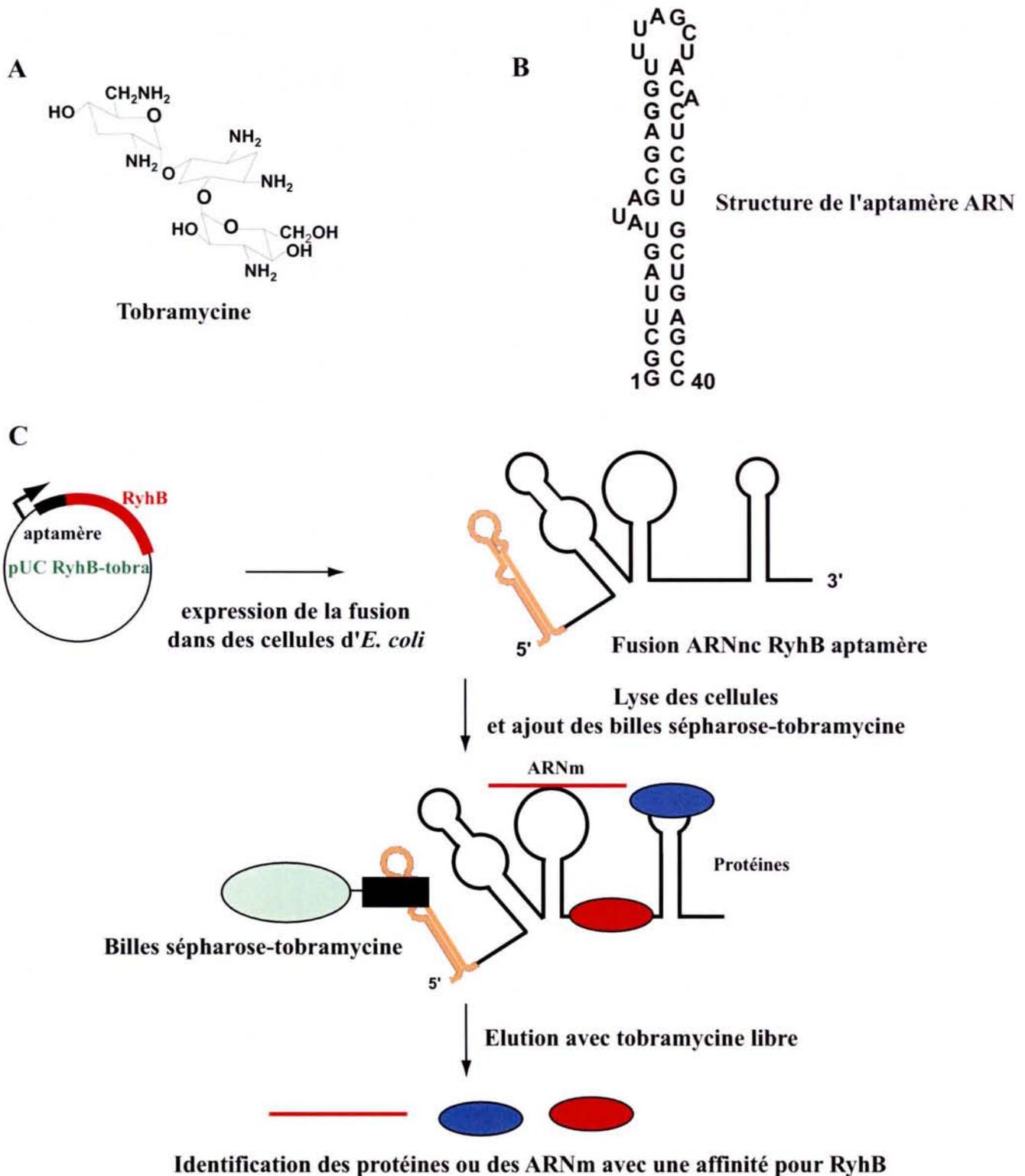


Figure 68 : Stratégie qui sera utilisée pour la recherche des partenaires de l'ARN RyhB *in vivo*.

A) Structure chimique de la tobramycine.

B) Structure 2D de l'aptamère ARN liant la tobramycine ($K_d = 5$ nM) (Cho *et al.*, 1998).

C) Etapes utilisées pour la purification des protéines et/ou d'ARN fixés sur l'ARN RyhB d'après la méthode mise au point par l'équipe de R. Lührmann (Hartmuth *et al.*, 2002) et développée au laboratoire dans le cadre de l'étude des rétrovirus. Les cellules d'*E. coli* sont transformées par la construction plasmidique permettant la production par transcription de l'ARN fusionné RyhB-aptamère. Ces cellules sont centrifugées puis lysées. Des billes de sépharose sur lesquelles la tobramycine a été greffée sont ajoutées dans le surnageant. Après centrifugation et lavage des billes, les molécules retenues sur l'ARN de fusion sont éluées avec de la tobramycine libre. Les protéines ou les ARN libérés sont ensuite identifiés par des techniques de caractérisation de protéines (spectrométrie de masse) ou d'ARN (RT-PCR et clonage).

yeaA par l'ARNnc RyhB permettrait de réguler la quantité de cette protéine MsrB stockant le fer en fonction de la concentration en ions Fe^{2+} présente dans le milieu de culture.

En conclusion, la région intergénique *gapA-yeaA* est complexe et un grand nombre de paramètres interviennent pour réguler l'expression des deux gènes. Ces différentes régulations interviennent au niveau de l'initiation de la transcription et pour cela, des facteurs de transcription généraux (CRP-AMPc et Fis) et l'existence d'une courbure de l'ADN sont mis en œuvre. Des régulations interviennent aussi sans doute au niveau traductionnel dans le cas du gène *yeaA* grâce à l'ARNnc RyhB.

Partie 3 : Etude de l'organisation transcriptionnelle du triplet d'ORF *epd-pgk-fbaA* et de la maturation des transcrits produits.

I.) Présentation de l'article.

Comme nous l'avons indiqué dans l'introduction, les gènes *gapA* et *pgk* codant la GAPDH et la phosphoglycérate kinase sont séparés physiquement sur le chromosome d'*E. coli* (Blattner *et al.*, 1997), alors que leurs produits agissent en succession dans la glycolyse. Ceci implique qu'ils soient corégulés. Comme nous l'avons déjà mentionné aussi, l'ORF localisée en amont du gène *pgk* chez *E. coli*, code une protéine très homologue aux GAPDH mais qui a une activité érythrose 4-phosphate déshydrogénase (Boschi-Muller *et al.*, 1997; Zhao *et al.*, 1995). L'idée est qu'au cours de l'évolution, dans l'un des ancêtres des γ -protéobactéries actuelles, le gène *gapA* ait été converti en un gène *epd*, ceci par un nombre limité de mutations. Alors que l'ORF *gapA* généralement localisée en amont de l'ORF *pgk* chez les bactéries est fortement exprimée, l'ORF *epd* est très faiblement exprimée (Boschi-Muller *et al.*, 1997; Zhao *et al.*, 1995). Enfin, dans les génomes de γ -protéobactéries, dont *E. coli*, l'ORF *pgk* est suivie par l'ORF *fbaA* codant une autre enzyme de la glycolyse, la fructose 1,6-bisphosphate aldolase (Alefounder *et al.*, 1989; Bardey, 2000).

V. Bardey, N. Robas et B. Charpentier avaient commencé à étudier le tandem d'ORF *epd-pgk*. Il ressortait de leurs travaux qu'un promoteur fort *gapB* P0, activé en présence de glucose, est présent en amont de l'ORF *epd* et permet la transcription des deux ORF *epd* et *pgk* (Charpentier *et al.*, 1998). V. Bardey et N. Robas avaient, par ailleurs, montré l'existence d'un promoteur en amont du gène *pgk* (*pgk* P1) et aussi observé l'existence d'un processus de clivage de l'ORF *epd* pouvant expliquer sa faible expression (Bardey, 2000). Enfin, V. Bardey et N. Robas avaient mis en évidence une faible efficacité de la région initiant la traduction (TIR) de l'ORF *epd* (Bardey, 2000).

La question se posait de savoir si l'ORF *fbaA* située en aval de l'ORF *pgk* était co-transcrite avec cette dernière, ou était transcrite de manière indépendante avec un promoteur qui lui était propre, comme cela avait été proposé (Salgado *et al.*, 2004). Pour cela, dans le cadre d'un travail de collaboration avec C. Vallet, étudiante en thèse, nous avons produit différentes

constructions renfermant les 3 ORF *epd-pgk-fbaA*, avec ou sans le promoteur *gapB* P0 et contenant un promoteur *pgk* P1 sauvage ou muté.

Une analyse par extension d'amorce des transcrits produits à partir de ces constructions a permis de montrer que les 2 promoteurs *gapB* P0 et *pgk* P1 sont responsables de la production de tous les transcrits obtenus à partir de la région *epd-pgk-fbaA*. Le promoteur *gapB* P0 conduit à la production de transcrits contenant les 3 ORF *epd-pgk-fbaA*, mais aussi de transcrits *epd-pgk*, lorsque le signal d'arrêt de la transcription rho-indépendant localisé en aval de *pgk* est utilisé. L'analyse par extension d'amorce des transcrits issus d'une construction renfermant les 3 ORF, confirmait que, comme V. Bardey et N. Robas l'avaient observé pour des constructions ne contenant que les ORF *epd* et *pgk*, les deux types de transcrits produits à partir du promoteur *gapB* P0 subissent des clivages à l'extrémité 5' de l'ORF *epd* et dans la région inter-ORF *epd-pgk*.

Le travail réalisé avec C. Vallet montre aussi que le promoteur *pgk* P1 permet la production de transcrit *pgk-fbaA* s'il n'y a pas d'arrêt au site de terminaison en aval de *pgk*, et *pgk* s'il y a terminaison à ce site. Les transcrits *epd-pgk-fbaA* et les transcrits *pgk-fbaA* sont tous deux clivés dans la région intergénique *pgk-fbaA*. En transformant différentes souches mutées dans les gènes des RNases (III, G et E) impliquées dans les processus de maturation et de dégradation des ARN, nous avons pu montrer que la RNase E est responsable des clivages à la fois dans l'ORF *epd*, la région intergénique *epd-pgk*, mais aussi dans la région intergénique *pgk-fbaA*. Par ailleurs, une étude de la stabilité relative du transcrit *pgk-fbaA* et des ARNm individuels *pgk* et *fbaA* obtenus au cours de ce processus complexe de transcription-maturation, a permis de montrer une plus forte stabilité des ARNm *fbaA* par rapport aux ARNm *pgk* et au tandem *pgk-fbaA*.

L'ensemble de cette étude montre que l'expression du triplet d'ORF *epd-pgk-fbaA* est régulée à plusieurs niveaux :

- 1) transcriptionnel, avec deux promoteurs, l'un régulé, *gapB* P0, fortement utilisé en présence de glucose, l'autre peu régulé, *pgk* P1, qui appartient à la classe des promoteurs étendus, mais qui, contrairement au promoteur *gapA* P1, ne nécessite pas une boîte -35 pour être actif ;
- 2) post-transcriptionnel avec l'existence d'un processus de clivage détruisant l'ORF *epd* mais aussi l'existence de clivages dans les régions inter-ORF qui permettent une segmentation des transcrits en segments correspondant à des ORF différentes, ces fragments ont des durées de vie différentes. Il apparaît de plus en plus que ce processus de segmentation en fragments ayant des durées de vie différentes est un moyen de réguler de manière différentielle l'expression des cistrons d'un même opéron.

3) traductionnel, avec en particulier une très faible efficacité de la traduction de l'ORF *epd*, qui semble liée à la fois à une distance sous-optimale entre la séquence RBS et le codon d'initiation, et à l'implication d'un facteur inhibiteur saturable agissant en *trans*. Grâce à ces multiples niveaux de régulations, un très faible taux de protéine E4PDH est produit, alors que les protéines PGK et FBA sont produites en grande quantité.

Ce travail a donné lieu à la rédaction d'un article soumis pour publication à la revue Journal of Bacteriology.

II.) Article II : Characterization of the molecular mechanisms involved in the differential production of Erythrose-4-phosphate dehydrogenase, Phosphoglycerate kinase and class II Fructose-1.6-bisphosphate aldolase in *Escherichia coli*.

V. Bardey, C. Vallet, N. Robas, B. Charpentier, B. Thouvenot, E. Hajndorf, P. Régnier, M. Springer and C. Branlant. Soumis à J. Bacteriol.

Publication n°2

Characterization of the molecular mechanisms involved in the differential production of Erythrose-4-phosphate dehydrogenase, 3-Phosphoglycerate kinase and class II Fructose-1,6-bisphosphate aldolase in *Escherichia coli*.

V. Bardey, C. Vallet, N. Robas, B. Charpentier, B. Thouvenot, E. Hajnsdorf, P. Régnier, M. Springer et C. Branlant.

Soumis à J. Bacteriol

Characterization of the molecular mechanisms involved in the differential production of Erythrose-4-phosphate dehydrogenase, 3-Phosphoglycerate kinase and class II Fructose-1,6-bisphosphate aldolase in *Escherichia coli*

V. Bardey^{1+#}, C. Vallet¹⁺, N. Robas¹, B. Charpentier¹, B. Thouvenot¹, E. Hajnsdorf², P.

Régnier², M. Springer², and C. Branlant^{1*}

+ The two authors contributed equally to this work.

(1) Laboratoire de Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Faculté des Sciences et Techniques, BP 239 , 54506 Vandoeuvre-lès-Nancy, Cedex, France, and (2) Institut de Biologie Physico-Chimique CNRS UPR 9073 conventionnée avec l'université Paris 7, Paris, France.

Running title: *epd-pgk-fbaA* expression in *E. coli*

* Corresponding author.

Mailing address : Maturation des ARN et Enzymologie Moléculaire, UMR CNRS 7567, Université H. Poincaré, Faculté des Sciences, 54506 Vandoeuvre-les Nancy, Cedex, France.

Phone: (33) 3 83 91 20 91

Fax: (33) 3 83 91 20 93

e-mail : Christiane.Branlant@maem.uhp-nancy.fr

present address: Laboratoires Sérobiologiques-Cognis France, 3 rue de Seichamps, BP40, 54 425 Pulnoy, France.

Abstract

A *gapA-pgk* tandem of genes coding the glyceraldehyde 3-phosphate dehydrogenase and 3-phosphoglycerate kinase, is frequently found in bacteria. However, in *enterobacteria* (e.g. *Escherichia coli*, *Salmonella typhimurium*, *Yersinia pestis* and *Vibrio cholerae*), *gapA* in this tandem was converted into an *epd* ORF coding an erythrose-4-phosphate dehydrogenase (E4PDH) and an *fbaA* ORF coding the class II fructose-1.6-bisphosphate aldolase follows *pgk*. Here, we show that in spite of its very low expression, the three *epd*, *pgk* and *fbaA* ORFs are cotranscribed from promoter *gapB* P0, that is activated in the presence of glucose, *pgk-fbaA* bicistronic RNAs were also found to be transcribed from an internal promoter of the extended -10 class (*pgk* P1). Analysis of *in vivo* RNAs revealed a segmentation process of both transcripts. It results in the degradation of the *epd* ORF and the production of stable *fbaA* mRNAs. RNase E was found to be responsible for this segmentation, and a cleavage site located within the *epd* ORF has the needed features to be an ectopic site. Its utilization is likely favored by the low translational efficiency of the *epd* ORF, that we demonstrated by use of *lacZ* translational and transcriptional fusions. This low efficiency is likely due to the short distance between the RBS and initiator AUG and the occurrence of a titrable translational inhibitor. Thus, *gapA* to *epd* conversion was likely accompanied by a decrease of the translational efficiency and the selection of an efficient RNase E ectopic site in order to limit *epd* expression.

Abbreviations : RT: reverse transcriptase;

G3P: glyceraldehyde 3-phosphate;

TIR: Translation Initiation Region;

RBS: Ribosome Binding Site

INTRODUCTION

E. coli is capable of using numerous carbohydrates as single source of carbon and energy. However, glucose is preferentially used in the presence of other carbon sources, so that glucose metabolism plays an important role for growth of *E. coli*. In all living cells, the phosphorylating glyceraldehyde 3-phosphate dehydrogenase (GAPDH), 3-phosphoglycerate kinase (PGK) and class II fructose-1,6-bisphosphate aldolase (FBA) are key enzymes for glucose metabolism (28). In most bacterial genomes, the gene encoding the phosphorylating GAPDH (*gap*) is located upstream of the PGK gene (*pgk*) (13, 15, 22, 25, 29, 53). In gram positive bacteria, such as *Bacillus megaterium* (70), *Bacillus subtilis* (48), *Corynebacterium glutamicum* (25), and *Lactobacillus delbrueckii* (15), the *gap-pgk* gene tandem is followed by a gene encoding another glycolytic enzyme, the triose phosphate isomerase (TPI). However, in γ -proteobacteria, that include several highly pathogenic bacteria strains such as *Escherichia coli*, *Salmonella typhimurium*, *Yersinia pestis*, *Pseudomonas aeruginosa*, *Shewanella putrefaciens*, and *Vibrio cholerae*, the *pgk* gene is followed by the *fbaA* gene encoding a class II fructose-1,6-bisphosphate aldolase and instead of being preceded by a *gap* gene, it is preceded by an *epd* gene (previously designated as *gapB*) (2, 6, 19, 23, 27, 78) (<http://www.tigr.org>). The *epd* gene encodes a protein showing strong sequence (11, 27) and 3D structure (Kauffman *et al.*, personal communication) similarities with GAPDHs. However, it has an erythrose-4-phosphate deshydrogenase activity (E4PDH) (11, 23, 78) and catalyzes the oxidation of erythrose-4-phosphate into 4-phosphoerythronate in the presence of NAD⁺ (78). Its biological role is still not completely understood, since E4PDH was proposed to be required for pyridoxal 5'-phosphate synthesis (77), however, strains with a deleted *epd* gene were found to be able to grow, although slowly, on minimum medium devoided of pyridoxal-5'phosphate (23). The *epd* gene of γ -proteobacteria is probably an interesting example of the conversion of a gene coding for a given dehydrogenase into a gene coding for another dehydrogenase. Indeed, it was proposed that the ancestor *gap* gene encoding a

phosphorylating dehydrogenase acting specifically on glyceraldehyde 3-phosphate (G3P) was at sometime during evolution converted into the present *epd* gene encoding a non-phosphorylating dehydrogenase, that acts efficiently on erythrose-4-phosphate (11). Conversion of a GAPDH into an E4PDH depends upon a limited number of amino acid substitutions. Four punctual amino acid replacements (D32E, T179M, P187Y, Y311C) and substitution of the SSTGA sequence around residue T208 in GAPDHs for the VDTKL sequence in E4PDHs seem to be responsible for the GAPDH to E4PDH conversion (11, 27). Residue T179 in GAPDHs is involved in binding of the G3P, so that the T179M substitution participates in the substrate specificity modification; T208 is implicated in binding of the inorganic phosphate, so that the VDTKL sequence found in E4PDH likely explains the absence of phosphorylating activity on erythrose-4-phosphate (11). However, as a relic of the previous activity, E4PDH has a residual poorly efficient phosphorylating dehydrogenase activity on G3P (11, 78). Nevertheless, it is not able to ensure growth on glucose in the absence of the *gapA* gene encoding the active phosphorylating GAPDH (23). The *gapA* gene, which may have arisen from a duplication of the ancestor gene or of an horizontal transfert of the genetic information (14, 52), is located at a distinct locus on the *E. coli* chromosome (40.1 minutes, instead of 66.2 minutes for the *epd-pgk* gene tandem) (2, 35). As the expression of the genes coding for glycolytic enzymes have to be coordinated, it is interesting to know how the regulations of the *gapA* and *pgk* glycolytic genes have evolved following their physical separation. It is also important to understand how the *epd* and *pgk* ORFs, that are transcribed from a common strong promoter *gapB* P0 (19, 67), can be differentially expressed. Indeed, in the presence of glucose, the *gapB* P0 promoter, which is located upstream of the *epd* ORF, has a strong activity and is responsible for about half of the cellular PGK activity (19). Nevertheless, in these conditions and other tested conditions, E4PDH is produced in very small amounts (2, 11, 78). Interestingly also, PGK and FBA are not present in equimolar

amounts in *E. coli* cells (59) and, up to now, the molecular basis of this differential expression was not studied.

Bacteria use several transcriptional and post-transcriptional mechanisms for the differential expression of proteins that are encoded by polycistronic mRNAs. The numerous previous studies in this field (for review (21, 66)) showed that internal transcription initiation, partial transcription termination, mRNA processing events, coupled with differential turnover of the maturation products, polarity and translational controls play important roles in determining the expression rates of individual cistrons within polycistronic transcripts. Surprisingly, in spite of the strong importance of glycolytic genes in *E. coli*, very limited information is available on the regulation of the expression of the *E. coli epd-pgk-fbaA* ORF triplet. We, thus, decided to determine which of the above mechanisms can explain the strong differential expression of the *epd*, *pgk*, *fbaA* ORF triplet. The first question to answer was to know whether the *epd-pgk-fbaA* cluster is an operon and/or if individual mRNAs are produced from internal transcription-start sites. Indeed, based on nucleotide sequence inspection, three transcription start-sites were previously predicted. However, only the presence of promoter *gapB* P0, its activation in the presence of glucose and its downregulation by the Cra protein in the presence of pyruvate were experimentally demonstrated (19, 64, 67). The fact that the *gapB* P0 promoter was found to be responsible for only half of the PGK production (19) and previous data obtained by Tn5 insertion mutagenesis (58) suggested the presence of an internal transcription-initiation site in the ORF cluster, which had to be identified. Finally, the activity of a predicted $E\sigma^{70}$ initiation site, located 30 nts upstream of the *fbaA* start codon (68), had not been tested experimentally. The possible activity of several putative rho-independent terminators, one located between the *pgk* and *fbaA* ORFs, and three other ones sitting, side by side, downstream from the *fbaA* ORF (2, 45) was also awaiting experimental support. In addition, as polycistronic mRNAs encoding GAPDH and PGK genes were found

to be subjected to processing mechanisms in *Zymomonas mobilis* and *Bacillus subtilis* (16, 48), the occurrence of such processing events in *E. coli* also had to be tested.

In the present study, by Northern-blot and primer-extension analyses, site-directed mutagenesis and S1 mapping, we established the transcriptional organization of the *E. coli epd-pgk-fbaA* genomic region, and studied the maturation and degradation processes of the transcripts. The RNase involved was identified by using *E. coli* strains defective in the production of various RNase(s). By construction of *epd-lacZ* transcriptional and translational fusions, we found a translational limitation of the *epd* expression in *E. coli*. Thus, as *V. cholerae*, which is responsible for the frequently fatal epidemic diarrheal disease cholera, is another γ -proteobacteria containing an *epd-pgk-fbaA* ORF cluster, (18, 34), we tested by study of the *V. cholerae epd* ORF whether translational limitation of the *epd* ORFs may be a general property of γ -proteobacteria. This was as much interesting to test, as the *V. cholerae epd* ORF is located downstream from the *irgA* gene encoding the major iron-regulated outer membrane protein, which is involved in virulence (18).

RESULTS

The *E. coli epd-pgk-fbaA* locus generates a complex array of mRNAs

To study the transcripts and cleavage products of the *epd-pgk-fbaA* ORF cluster (Fig. 1A), this ORF cluster was cloned into a multicopy plasmid, together with the 243-bp long and 431-bp long upstream and downstream DNA regions, respectively (plasmid pEPF) (Fig. 1 B). A shorter construct, missing promoter *gapB* P0, was built (plasmid p Δ P0-EPF) (Fig. 1B). The *E. coli* TG1 strain transformed with plasmids pEPF or p Δ P0-EPF (designated as pEPF or p Δ P0-EPF transformed cells) or with the empty pBS plasmid (control transformed cells) were grown until mid-log phase in M63 minimal medium, in the presence of glucose as the sole source of carbon. Total RNA extracted from the transformed cells were analyzed by Northern-blot, using oligonucleotides (O-EPD, O-PGK, and O-FBA), that are complementary to each of the three *epd*, *pgk* and *fbaA* ORFs, respectively (Fig. 1A). In figure 2, for each transformed strain, total RNA was extracted from a similar number of cells (about 1.5×10^9). Then, the fraction of total RNA, that was loaded on the agarose-formaldehyde gel, was inversely proportional to the plasmid copy number of the transformed cells. Indeed, plasmid p Δ P0-EPF was found to be present at a higher copy number compared to plasmid pEPF (factor of 2 to 10 depending on the experiment). The experiment was repeated several times, and reproducibly, each of the oligonucleotide revealed mRNAs ranging in size from 3.7 to 1.1 kbs.

For plasmid pEPF, a very faint band containing an RNA with an estimated length of about 3.7 kb was revealed with the 3 probes. As it was absent in both the control cells and the p Δ P0-EPF transformed cells, it corresponds to the tricistronic *epd-pgk-fbaA* transcripts, which were initiated at promoter *gapB* P0. For plasmid pEPF, each of the 3 probes also detected RNAs, that were expected to contain 2 ORFs according to their sizes (2.4 to 2.7 kbs). However, the O-EPD probe gave a very faint signal, compared to the two other probes. This faint signal is

expected to correspond to an *epd-pgk* bicistronic RNA initiated at promoter *gapB* P0. It might result either from a transcription arrest at the putative terminator located downstream from the *pgk* ORF, or from cleavage of the *epd-pgk-fbaA* transcript by RNase(s). The detection with the O-PGK and O-FBA probes of bicistronic transcripts in cells that were transformed with plasmid pΔPO-EPF, which is missing promoter *gapB* P0, was a strong indication for the presence of an internal transcription-start site directing the expression of the *pgk* and *fbaA* ORFs.

In the pEPF transformed cells, the 3 probes also enlightened RNAs that were expected to be monocistronic according to their sizes (1.1 to 1.5 kbs). The absence of putative transcriptional terminator between the *epd* and *pgk* ORFs indicated that the detected 1.2 kb *epd* transcript resulted from the cleavage of a longer molecule. As expected, after promoter *gapB* P0 deletion, no *epd* monocistronic RNA was detected with the O-EPD probe (Fig. 2).

Altogether, the data suggested the presence of at least one internal transcription-start site, located upstream of the *pgk* ORF and the occurrence of either a second transcription-start site located upstream of the *fbaA* ORF or a cleavage process. The broad bands found for bi- and mono-cistronic RNAs and the presence of small size RNAs in total RNA extracted from the pEPF transformed cells suggested a complex degradation or maturation process of the *gapB* P0 transcript.

A promoter of the extended -10 class is located upstream of the *pgk* ORF

To identify the transcription promoter located upstream of the *pgk* ORF, the *E. coli* TG1 strain was transformed with plasmids pPBK100 and pBK200, containing the *epd* and *pgk* ORFs with or without promoter *gapB* P0, respectively (Fig. 1B). These plasmids, with smaller sizes, were used because of their stronger stability in *E. coli* compared to plasmid pEPF. Total RNA were extracted from transformed cells grown on glucose. The 5' extremities of the PGK mRNAs were identified by reverse transcription using oligonucleotide O-PGK as the primer

and a sequencing ladder generated by the same oligonucleotide (Fig. 3A₁). The occurrence of transcription initiation at promoter *gapB* P0 was verified by extension of the O-EPD oligonucleotide with RT (Fig. 3A₂). In parallel, the 5'-end labeled O-GB1/O-PGK DNA probe (Fig. 1B), described in Materials and Methods, was used to identify the 5' extremities of the PGK mRNAs by S1 nuclease mapping (Fig. 3B). Both primer extension analysis and S1 mapping experiments were repeated several times and reproducible results were obtained.

As illustrated in Figure 3A, in the presence of promoter *gapB* P0, several RT stops were detected between positions +967 (in the *epd* ORF) and +1184 (in the *epd-pgk* intergenic region). In contrast, no strong RT stop was detected in the 400-nt region located upstream of position +967 (not shown). Almost all the RT products obtained in high yields had a counterpart among the nuclease S1-mapping products, indicating that they corresponded to RNA extremities. The same *pgk* mRNA 5' extremities were detected by RT analysis of total RNA extracted from untransformed cells (Fig. 3B₁, lane 1). Hence, at least, six distinct *pgk* mRNAs, differing by the identity of their 5' extremities, were present in total RNA extracted from both transformed and untransformed cells. They were designated as *pgk* mRNAs 1 to 6 and the location of their 5' extremities in the *epd-pgk* sequence is shown in figure 3D.

Upon deletion of promoter *gapB* P0, all the cDNA bands had a strongly decreased intensity, except that at position +967 (Fig. 3A₁, compare lanes 1 and 4). We concluded that the presence of the *pgk* 1 mRNA was not dependent on the activity of promoter *gapB* P0. Accordingly, a TATAGT sequence, carrying the characteristic features of the -10 elements of promoters recognized by the $E\sigma^{70}$ RNA-polymerase holoenzyme (33, 47), is located at the expected 7-bp distance from position +967 (Fig. 3D). No obvious -35 element is detected. However, as the TATAGT sequence is preceded by a TGG triplet, the corresponding promoter may belong to the special class of *E. coli* extended -10 promoters, which may function without -35 hexamer (12, 56). As a confirmation, the replacement of the putative -10 hexamer segment (TATAGT, position +954 to +959) for a CACAGC sequence (plasmid

pmut-10P1-PBK100) (Fig. 3A, lane 3) and the substitution of the TG dinucleotide (positions +951 to +952) for an AC dinucleotide (plasmid pmutTGP1-PBK100) (Fig. 3A, lane 2) led to the complete disappearance of the RT stop at position +967 (Fig. 3A, lanes 2 and 3, respectively). This demonstrated the presence of a promoter with an extended -10 sequence upstream of the *pgk* ORF. It was designated promoter *pgk* P1.

Both PGK and FBA productions depend upon the *gapB* P0 and *pgk* P1 transcription start sites

To test whether a third transcriptional promoter is present upstream of the *fbaA* ORF, we built plasmid p Δ P0-mut-10P1-EPF, that carried both the *gapB* P0 deletion and the CACAGC sequence instead of the WT-10 sequence of promoter *pgk* P1 (Fig. 1B). In accord with the above data, no bicistronic *pgk-fbaA* mRNAs and no monocistronic *pgk* and *fbaA* mRNAs were detected by Northern-blot analysis of total RNA extracted from cells transformed with this plasmid (Fig. 2). Thus, individual *fbaA* mRNAs are produced from transcripts initiated at the *gapB* P0 and/or *pgk* P1 promoters. PGK and FBA activity measurements confirmed this hypothesis. Indeed, in cells transformed with the p Δ P0-mut-10P1-EPF plasmid, the PGK and FBA activities were reduced to levels corresponding to the yields of expression of the endogenous genes (same levels as with the empty pBS plasmid) (Table 3). Hence, only promoters *gapB* P0 and *pgk* P1 contribute to the PGK and FBA productions. In spite of the presence of a putative specific *fbaA* promoter upstream of the *fbaA* ORF (68) and a putative rho-independent terminator downstream from the *pgk* ORF (Fig. 1) (2), the *fbaA* ORF is predominantly cotranscribed with the *pgk* ORF. Taking into account both the PGK and FBA activities measured in extracts of the pEPF and p Δ P0-EPF transformed cells and the difference of plasmid copy numbers in these transformed cells, promoters *gapB* P0 and *pgk*

P1 are expected to have similar contributions in PGK and FBA productions, when glucose is present in the medium (Table 3).

To test for the influence of the presence of glucose in the medium on the activity of promoter *pgk* P1, untransformed TG1 cells were grown in M63 medium supplemented with glucose or pyruvate. RT analysis of the PGK mRNAs in total RNA extracted from cells grown on these two carbon sources revealed no significant variation of the *pgk* P1 transcript concentration as a function of the carbon source (Fig. 3C). We concluded that promoter *pgk* P1 is not activated in the presence of glucose in the medium. In accord with the data obtained after promoter *gapB* P0 deletion, the low activity of promoter *gapB* P0 in the presence of pyruvate (Fig. 3C₂) was accompanied by an almost complete disappearance of the *pgk* mRNAs 2 to 6 (Fig. 3C₁). This strongly suggested that these *pgk* mRNAs are generated by cleavage of the *epd-pgk* and/or *epd-pgk-fbaA* transcript(s), but not by cleavage of transcripts initiated at the *pgk* P1 promoter. In accordance with the idea that they correspond to cleavage products, none of the identified 5' extremities of the *pgk* mRNAs 2 to 6, are preceded by a putative promoter sequence recognized by one of the *E. coli* RNA-polymerase holoenzymes (for review,(76)). It should be noticed, that whereas the cleavage sites 5 and 6 are located within the *epd-pgk* inter-ORF region, sites 2, 3 and 4 are located in the 3'-terminal part of the *epd* ORF (Fig. 3D). Consequently, cleavages at these 3 sites are expected to abolish E4PDH production. In addition, whereas sites 2 to 6 are present in the *pgk* P1 transcripts, as mentioned above, this transcript is inefficiently cleaved compared to the *gapB* P0 transcript.

Individual *fbaA* mRNAs are produced by maturation of the *epd-pgk-fbaA* and *pgk-fbaA* transcripts

To verify that the *fbaA* individual mRNAs, that we detected by Northern-blot analysis (Fig. 2) were generated by cleavage of the *epd-pgk-fbaA* and/or *pgk-fbaA* transcripts, and in order to identify the cleavage sites, the *fbaA* mRNAs contained in total RNA extracted from the pEPF

or p Δ P0-EPF transformed cells were analyzed by extension of primer O-FBA with RT (Fig. 4, A and B). None of the detected RT stops corresponded to the previously predicted transcription-start site (68) (not shown). Successive RT stops were found in the series of adenosine residues from position +2364 to +2369 (*fbaA* 5 series), and at four other positions denoted *fbaA* 1 to 4 (Fig. 4, panel A and D). Three of these stops (*fbaA* 1, 3 and 4) were also detected by RT analysis of the F_{*pgk-fbaA*} *in vitro* transcript (position +2265 to +2626) with the O-FBA probe (Fig. 1B) (Fig. 4B, lane 3). Hence, the RT stops designated *fbaA* 1, 3 and 4, in the analysis of cellular RNAs were likely due to RT pauses. Only the stop at the *fbaA* site 2 and the *fbaA* 5 series of stops corresponded to *in vivo* cleavage sites. Accordingly, no clearly identified transcription-promoter was detected upstream of these RT stops (Fig. 4D). The *fbaA* 5 series of sites and site *fbaA* 2 are located on both sides of the potential *pgk* transcription terminator, respectively (Fig. 4D). As cleavages from position +2364 to +2369 overlap the *pgk* UAA terminator codon of the *pgk* ORF, the *pgk* cleavage products are expected to be unstable. An identical pattern of cleavages was detected in RNAs from cells transformed with the p Δ P0-EPF plasmid, indicating that the cleavage process also takes place in the *pgk-fbaA* transcript (Fig. 4B, lane2).

The *epd* ORF and the *epd-pgk* and *pgk-fbaA* inter-ORF regions are cleaved by RNase E *in vivo*

Three endonucleases were found to be involved in mRNA degradation and maturation in *E. coli*: RNase III, RNase E (for review(30) and the more recently identified RNase G (for review, (44, 75). To test for a possible implication of RNase III in cleavage of the *epd-pgk*, *pgk-fbaA* and *epd-pgk-fbaA* transcripts, we transformed the two *E. coli* isogenic strains IBPC602 and IBPC600 (Table 1) with plasmid pEPF. The presence of the *rnc-105* mutation in strain IBPC600 abolishes RNase III activity. Total RNA from the two transformed strains were studied by primer-extension analysis. The patterns of cDNAs obtained with both the O-

PGK and O-FBA primers were identical for the RNase III⁻ and the wild type strains (data not shown). When similar experiments were performed with the N3433 strain and its isogenic strain IBPC935 containing a disrupted RNase G gene (Hajnsdorf *et al.*, unpublished data) (Table 1), again, similar patterns of cDNAs were obtained for both oligonucleotides (data not shown). Thus, we concluded that the RNases III and G are not responsible for cleavages in the *epd-pgk* and *pgk-fbaA* inter-ORF regions.

We thus tested the possible implication of RNase E, by using the MG1693 and SK5665 isogenic strains (3) (Table 1). Due to the *ams-1* mutation, strain SK5665 (RNase E^{ts}) produces a thermosensitive RNase E. As RNase E is essential for growth of *E. coli*, the two transformed strains, were grown at 30°C (permissive temperature) until mid-log phase. Then, half of each cultures was subjected to a temperature shift at 42°C. RT analysis of the *pgk* mRNAs with the O-PGK probe showed a disappearance of the *pgk* mRNAs 2 to 6 after incubation of the RNase E^{ts} strain for 15 min at 42°C (Fig. 5A). Such disappearance was not observed for the WT strain. By using primer O-EPD, we verified that this disappearance was not due to the absence of transcription initiation at promoter *gapB* P0 (not shown). Based on these observations, we concluded that RNase E is responsible for the cleavages that generate the *pgk* mRNAs 2 to 6.

RT analysis of the *fbaA* mRNAs from the WT and RNase E^{ts} strain with the O-FBA probe (Fig. 4C) showed that cleavage in the series of A residue from position +2364 to +2369 is temperature sensitive in the WT strain. However, a stronger decrease of the cleavages in the *fbaA* 5 series of sites and at site *fbaA* 2 was observed in the RNase E^{ts} strain. We, thus, concluded that RNase E is also responsible for the cleavages detected in the *pgk-fbaA* interORF region.

Similar patterns of *epd-pgk* cleavages by RNase E are obtained *in vitro* and *in vivo*

To confirm the implication of RNase E in production of the *pgk* mRNAs 2 to 6, we tested the activity of a degradosome fraction of the *E. coli* RNase E^{ts} strain on the F_{*epd-pgk*} RNA (positions +701 to +1500) (Fig. 1), produced by *in vitro* transcription. As RNase E is highly dependent on RNA secondary structure, the F_{*epd-pgk*} RNA was designed in order that 292- and 316-nt long segments were present 5' and 3' to the run of cleavage sites, respectively. The possibility to abolish the RNase E activity of the degradosome by a 10 min incubation at 45 °C (3) was used as an internal control. The *in vitro* maturation products were analyzed by reverse transcription, using oligonucleotide O-PGK as the primer. The extension products were compared with those obtained for total RNA extracted from the MG1693 (WT) strain transformed with plasmid pPBK100. When the *in vivo* and *in vitro* data were compared, we noticed a slight shift by 1 or 2 nts of the cleavage sites and sometimes, the presence of two or three cleavages in succession instead of one single cleavage in the *in vitro* experiment compared to the *in vivo* assay (Fig. 5, panel B and C). Nevertheless, very similar patterns of extension products were obtained for both experiments. In addition, the RT stops common to the *in vitro* and *in vivo* assays were absent in the control experiment (preincubated degradosome fraction). Altogether, the data strongly reinforced the idea that RNase E is directly responsible for the observed cleavages.

The *fbaA* mRNA has a strong stability *in vivo*

Inhibition of transcription by rifampicin was used to study the relative stabilities of the RNAs containing the *pgk* and *fbaA* ORFs. To this end, the *E. coli* MG1693 (WT) strain, transformed with plasmid pEPF, was grown at 30°C until mid-log phase in LB medium. Then, rifampicin was added to the cell culture at a 0.5 mg/ml concentration and growth was continued. Samples were collected before, at the time, and at different times after rifampicin addition, and Northern-blot analyses of total RNA were performed with the O-PGK and O-FBA probes. As

in LB medium, promoter *gapB* P0 is poorly used, total RNA essentially contained the bicistronic *pgk-fbaA* and monocistronic *pgk* and *fbaA* RNAs (Fig. 6). Thus, we studied the relative stabilities of these 3 mRNAs. By measurement of the radioactivity in the Northern-blot signals, the half-life of the bicistronic RNAs was estimated to be of about 4 ± 0.4 min (Fig. 6). The monocistronic *pgk* RNA was expected to correspond to primary transcripts initiated at promoter *pgk* P1 and terminated at the *pgk* transcription terminator, together with cleavage products of larger RNAs. The monocistronic *fbaA* RNA mainly arised from cleavage of the *pgk-fbaA* transcripts. Consequently, the half-lives, that we estimated directly from measurement of the *pgk* and *fbaA* RNA signals (4.5 ± 0.3 and 12 ± 1.5 min, respectively) (Fig. 6), were apparent half-lives. This difference of apparent half-life was in agreement with the fact that *pgk* monocistronic RNA cleaved close to or within the UAA terminator codon are expected to be unstable. As *fbaA* monocistronic RNAs were digestion products of the bicistronic RNAs, we could estimate their real half-life by using the mathematical model given in Materials and Methods and assuming that the disappearance of the *pgk-fbaA* bicistronic RNA was essentially due to its cleavage into *pgk* and *fbaA* individual mRNAs. The *fbaA* mRNA real half-life was estimated to be of 6 min, which shows that it is the most stable of the RNAs produced from the *epd-pgk-fbaA* ORF cluster.

Cleavage of the *epd* ORF is probably due to its low translation efficiency

As previously observed (36), the occurrence of RNase cleavages inside the *epd* ORFs might reflect a low translation efficiency. We built *epd-lacZ* transcriptional and translational fusions to test whether the *epd* ORF has a translation-initiation region (TIR) of low efficiency. In plasmid pVBGF2 (*epd-lacZ* translational fusion), a β -galactosidase truncated of its 6 N-terminal amino acids and fused to the 18 N-terminal amino acids of E4PDH, was expressed under the control of promoter *gapB* P0 and the *epd* TIR. In plasmid pVBOF2 (*epd-lacZ* transcriptional fusion), the WT β -galactosidase was expressed under the control of promoter

gapB P0 and the *lacZ* TIR (Fig. 8, panel A and C). The cellular β -galactosidase-specific activity was 12-fold higher for cells transformed with the transcriptional fusion compared to cells expressing the translational fusion (Table 4). By restoration of an entire β -galactosidase ORF in plasmid pVBGF2, only a limited increase of the β -galactosidase was observed (factor of 2) (data not shown). Hence, these results were an indication for a limited efficiency of the *epd* TIR. To confirm this hypothesis, we used the λ RS45 phage to introduce the transcriptional and translational fusions as unique copies in the *E. coli* TG1 chromosome. The cellular β -galactosidase-specific activity of the lysogenic strain containing the translational fusion was 150 fold lower, compared to that of the lysogenic strain containing the transcriptional fusion (Table 4). This brought strong additional support to the idea that the *epd* TIR has a low efficiency. In addition, the stronger difference observed between the transcriptional and translational fusions, in the monolysogenic cells compared to the transformed cells, suggested the presence of a translational inhibitor of the *epd* ORF, which is titrated by the large *epd* ORF copy number in the transformed cells.

The small distance between the RBS and initiator AUG probably limits the *epd* TIR efficiency

The *epd* TIR has two peculiar features that may explain its low efficiency: first, a long stem-loop structure can be formed immediately upstream of the Ribosome Binding Sequence (RBS) (Fig 7D) and, secondly, the distance between the RBS and AUG initiation codon is unusually short (aligned distance of 5, instead of the 8-10 nts optimal distance) (49) (Fig. 7E). To test whether these two features may be involved in translation limitation, two variant *epd* TIRs were produced. In variant 35, the 5' strand of the stem (29 nts) was substituted for a GTCGAC sequence, so that formation of the long stem-loop structure was abolished. In variant 408, the WT AUCU sequence, that links the RBS to the AUG, was replaced by an 8-nt

sequence (AAAAUAAA) (aligned distance of the mutated RBS of 9 nts) (Fig. 7D). Each of these two mutations was introduced in the pVBGF2 and pVBOF2 plasmids, described above, and were integrated as unique copies in the chromosome of the *E. coli* TG1 strain. Measurement of the β -galactosidase specific activities of the transformed cells and monolysogenic cells (Table 4) showed that both TIR mutations had no marked effect on the expression of the transcriptional fusions. Mutation 35 also had a very limited effect on the translational fusion. In contrast, introduction of an optimal spacing between the RBS and initiation codon (variant 408) increased the β -galactosidase activity by a factor of about 5 for the transformed cells and 15 for the monolysogenic cells. Hence, the small distance between the RBS sequence and the AUG initiation codon probably contributes to the low activity of the *epd* TIR.

Mutations 35 and 408 were also introduced in plasmid pPBK500 containing the *epd-pgk* gene tandem (Fig. 1B and Table 5). Alteration of the stem-loop structure had no marked effect on both the PGK and E4PDH activities. In contrast, increasing the distance between the RBS sequence and initiator AUG led to the doubling of the E4PDH activity, whereas the PGK specific activity was unmodified. This reinforced the idea that the *epd* TIR limits *epd* translation efficiency.

Expression of the *V. cholerae epd* ORF is likely translationally limited

As mentioned in the introduction, an *epd* ORF was also detected upstream of the *pgk* gene in the *V. cholerae* genome (18). It was interesting to know whether the same mechanism was used to limit *epd* compared to *pgk* expression in this bacteria species. Inspection of the sequence upstream of the *V. cholerae epd* ORF also revealed the presence of a short distance between the RBS sequence and the initiation codon (distance of 3 nts, aligned distance of 6 nts) (Fig. 7E). As above, to test for the efficiency of the *V. cholerae epd* TIR, we built translational and transcriptional fusions using the *V. cholerae epd* or *pgk* gene and the *E. coli*

lacZ gene (Fig. 7, panel B and C). The four fusions were studied in multicopy plasmids. As shown in Table 6, similar β -galactosidase activities were obtained for the transcriptional and translational *pgk-lacZ* fusions. This suggests that the *V. cholerae pgk* TIR is at least as much efficient as the *E. coli lacZ* TIR. As found for the *E. coli epd* TIR, the β -galactosidase activity measured for the *epd-lacZ* translational fusion was about 1/10 of that measured for the *epd-lacZ* transcriptional fusion, suggesting a low efficiency of the *V. cholerae epd* TIR due to the short distance between the RBS and AUG initiator codon. Accordingly, the E4PDH activities produced in *E. coli* cells transformed with plasmid p4Vc27 or pPBK500 (Table 5), that were, respectively, expressing the *V. cholerae* or *E. coli* E4PDHs under their own transcription and translation signals, were very similar ($0.47 \pm 0.15 \mu\text{mol NADH}/\text{min}/A_{280}/A_{600}$ against $0.46 \pm 0.04 \mu\text{mol NADH}/\text{min}/A_{280}/A_{600}$, respectively).

DISCUSSION

The *pgk* and *fbaA* genes are two of the predicted more highly expressed genes in *E. coli* (40) and were experimentally demonstrated to be among the more highly expressed genes in *E. coli* K12 grown on glucose, acetate, gluconate or glycerol (59). The *E. coli epd* ORF is expressed at very low level (11, 78). However, here we show that the *epd*, *pgk* and *fbaA* ORFs are cotranscribed in *E. coli*. By a deep analysis of the expression of the *E. coli epd-pgk-fbaA* operon, we showed that translational limitation and RNase E cleavage of the *epd* ORF both participate to the very low expression of this ORF. The second RNase E cleavage region, that we detected in the *pgk-fbaA* interORF region is likely involved in the differential expression of the *pgk* and *fbaA* ORFs. Altogether, our data demonstrate the occurrence of several levels of regulation of the *E. coli epd-pgk-fbaA* operon, including transcription, translation and RNA cleavage processes.

We found a single internal promoter (*pgk* P1) in the *E. coli epd-pgk-fbaA* operon. In the various fermentation conditions, that we tested [M63 medium supplemented with glucose or pyruvate, and LB medium (not shown)], no specific *fbaA* promoter was used, so that the production of individual *fbaA* mRNAs was dependent upon RNase E cleavage of transcripts initiated at the *gapB* P0 and *pgk* P1 promoters. Promoter *pgk* P1 belongs to the class of extended -10 promoter. In early studies on *E. coli* promoters of this class, the additional TG dinucleotide, that is present one base pair upstream from the -10 hexamer was proposed to replace functionally the -35 hexamer (41). Indeed, this dinucleotide interacts with the $\sigma^{3.0}$ domain of factor σ^{70} (7, 41, 43). However, a more extensive study on *E. coli* promoters showed that promoters with an extended -10 sequence may also have a functional -35 hexamer (56). We, recently, brought experimental support to this assumption, by studying the *E. coli gapA* P1 promoter, which directs GAPDH production in *E. coli* (74). However, in contrast to promoter *gapA* P1, the -35 sequence of promoter *pgk* P1 shows no significant homology with the consensus sequence of -35 hexamers (2 matches on 6 bp). Hence, we

propose, that promoter *pgk* P1 does not need a -35 hexamer for activity. The need for a greater complexity of promoter *gapA* P1 may be explained by its strong activation in the presence of glucose (19) or an energy source aminoacid in the medium (31). Interestingly, expression of the *E. coli gapA*, *pgk* and *fbaA* ORFs coding for three enzymes, which acts in succession in glycolysis, depends upon promoters of the extended -10 class. Promoters of this class seems to be frequently used for genes coding for glycolytic enzymes since the *pfkB* (phosphofructokinase) and *pckA* (phosphoenolpyruvate carboxykinase) genes are also under the control of a promoter of this class (56, 63). Promoters of the extended -10 class without -35 hexamer are more easily accommodated in ORFs, as is the case for promoter *pgk* P1 and are expected to ensure a basal expression of genes independently of the growth conditions. Accordingly, we found that the amount of transcript, initiated at promoter *pgk* P1 is very similar in the presence of pyruvate or glucose in the medium. Altogether, our data suggest that promoter *pgk* P1 plays a very important role for PGK and FBA productions in the absence of glucose, whereas both promoters *gapB* P0 and *pgk* P1 are strongly implicated in the production of these two enzymes in the presence of glucose.

The *E. coli* RNases E, G, III and P were found to be involved in transcript processing and mRNA decay (30, 44, 46). Here, we found that RNase E plays a key role in segmentation of transcripts from the *epd-pgk-fbaA* ORF cluster. RNase E is a single stranded specific RNase, that acts within A and/or U rich segments (24, 38) and cleaves RNAs with a 5' monophosphate group in preference to RNAs which are triphosphorylated at their 5'-end (50). Based on its higher activity on A and/or U rich sequence, RNase E would be expected to have a general role in the initiation of mRNA decay, that is essentially dependent on the length of the mRNAs. In fact, this is not the case (5), and the expected correlation between translational efficiency and mRNA stability (36) cannot explain the hierarchy of *E. coli* mRNA sensitivity to RNase E (5). Starting from these statements, the notion of ectopic (or internal) RNase E sites has been proposed (5, 51) and our data bring strong support to this proposal. In addition,

by taking into account our data and recent investigations on the RNase E specificity (38, 39, 65), we can try to figure out the characteristic features of the RNase E ectopic sites.

Indeed, we noticed that the *pgk 2* site, which is the most upstream site in the *pgk 2* to 6 cluster of RNase E sites, and that may thus be an ectopic site, has sequence and structural properties that were proposed to be important for the docking of RNase E on its substrate and for RNase E catalytic efficiency. An experimental study of the secondary structure of the $F_{epd-pgk}$ transcript (V. Bardey, unpublished data) showed that site *pgk 2* is located 2 nts downstream from a G residue in an A/U rich single-stranded segment, which is flanked by two stable stem-loop structures (Fig 8B₁). The presence of a G residue 2 nts upstream of the scissile bond was found to have a strong importance for the catalytic efficiency of RNase E (38, 65). Among the 5 sites in the *epd-pgk* region, only site *pgk 5* has no G residue 1 or 2 nts upstream of the scissile bond (Fig. 3 and 5). The presence of two stable stem-loop structures flanking site *pgk 2* is not shared by the *pgk* sites 3 to 6 that are found in a poorly structured RNA region. This peculiar 2D structure of site *pgk 2* should be put in light of previous observations showing that the presence of stem-loop structures 5' and 3' to the scissile bond influences the rate of RNase E cleavage (for review, (30)) and may be required for the loading of RNase E onto its substrate (39). Altogether, this suggests that site *pgk 2* within the *epd* ORF has all the needed characteristic features to be efficiently recognized by RNase E as an internal entry site. Binding of RNase E to this site may be mediated by association of the 5' and/or 3' stem-loop structures with the ARRBD C-terminal domain of RNase E (39), thus avoiding the requirement for a 5' monophosphate. In the *pgk P1* transcript, site *pgk 2* is not preceded by a stable 5' stem-loop structure (Fig. 8B₁). This may abolish its capacity of docking on RNase E, and thus explain the inefficient cleavage of the *pgk P1* transcript.

Hence, we propose the following model for the processing of the *epd-pgk* and *epd-pgk-fbaA* transcripts (Fig. 8). Site *pgk 2* is an RNase E internal entry (or ectopic site) site. Depending on the identity of the cleaved transcript, cleavage at this site generates a 3' product containing the

pgk ORF or both the *pgk* and *fbaA* ORFs. This 3' product may then be subjected to further cleavages at sites *pgk* 3 to 6 by RNase E, because after initial cleavage at site *pgk* 2, the 5' monophosphate of the 3' product bound to the enzyme increases its catalytic activity (37), and in these conditions, sites which have not the needed properties to be ectopic sites may be cleaved. We can assume that ribosomes translating the *pgk* and *fbaA* ORFs constitute barriers that limit RNase E action to the *epd-pgk* and *pgk-fbaA* ORF regions. In contrast, due to the low translation efficiency of the *epd* ORF, the 5' product released by RNase E cleavage at site *pgk* 2 is likely rapidly degraded by action of the degradosome. However, by Northern blot and S1 analyses, we detected the presence of a 484-nt long resistant fragment corresponding to the 5' terminal region of the transcript (data not shown). This resistant fragment carries stable stem-loop structures at both extremities: the one at the 5' extremity corresponds to the long stem-loop structure of the 5' UTR region (Fig. 7D) and that at the 3' extremity is an internal stem-loop structure of the *epd* ORF (Fig. 8C). These two stable structures probably explain the observed resistance to nucleases.

As the 5' UTR of the *pgk* P1 transcripts is inefficiently cleaved by RNase E, cleavage of these transcripts in the *pgk-fbaA* interORF region is expected to involve at least one other internal RNase E internal entry site. An experimental study of the secondary structure of the *pgk-fbaA* interORF region performed on the $F_{pgk-fbaA}$ transcript (C. Vallet, unpublished results) shows that both the *fbaA* 5 series of sites and site *fbaA* 2 are located in an A/U rich single stranded segments flanked by stable helical regions (Fig. 8B₂). The presence of a G residue 2 nts upstream of the scissile bond is only found for the more upstream of the *fbaA* 5 sites.

As already mentioned, the *pgk* individual mRNAs resulting from cleavages in the series of *fbaA* 5 sites are missing the 3' terminal stem-loop structure formed by the transcription terminator. In addition, the *fbaA* 5 cleavages are located very close or within the UAA termination codon, so that *pgk* mRNAs cleaved at these sites are not expected to be stable. Hence, stable individual *pgk* mRNAs are expected to correspond to transcripts cleaved at site

fbaA 2 and to *pgk* P1 or *gapB* P0 transcripts stopped at the *pgk* P1 transcription-terminator signal.

In contrast, the *fbaA* individual mRNAs resulting from the segmentation of longer molecules were found to be stable (real half-life of 6 min), which may be explained by the presence of stable stem-loop structures at the two extremities of *fbaA* mRNAs. Indeed, cleavages at sites *fbaA* 5 or *fbaA* 2 both release a 3' product with a stable 5' stem-loop structure (Fig. 8). In addition, the *fbaA* ORF is followed by Repeated Extragenic Palindromic sequence (REP) of the BIME 2 class (4) (Fig. 8D), that forms three successive putative rho-independent terminators (Fig. 1A and 8D), and a rho-independent terminator (26). They likely constitute a barrier against exonuclease action.

The occurrence of processing events leading to the production of individual *pgk* mRNAs was also found in bacteria containing a *gapA-pgk* tandem instead of an *epd-pgk* tandem. However, the RNase(s) involved in the cleavage processes have never been identified. In *Zymomonas mobilis*, cleavage into individual *gapA* and *pgk* mRNAs of the *gap-pgk* transcript leads to highly stable *gapA* mRNAs and less stable *pgk* mRNAs (16). In *Lactobacillus delbrueckii*, individual *gapA* and *pgk* mRNAs were proposed to be generated by cleavage of the *gap-pgk-tpi* transcript. Here also, the individual *gapA* mRNA is more stable than the *pgk* mRNA (15). More recently, the complex transcriptional organization of the *gap-pgk-tpi-pgm-eno* ORF cluster of *Bacillus subtilis* was elucidated (48). Strong promoters are found upstream of the *gapA* and *pgk* ORFs, and the *gapA* ORF can be cotranscribed with the ORF of the transcriptional regulator CggR. This bicistronic transcript is processed by an RNase, which has a specificity similar to that of *E. coli* RNase E (54). Taken together, these data and our data are in favour of a strong importance of transcript segmentation for modulation of the expression of the glycolytic genes.

Based on our study of the *E. coli epd* TIR, the consequences of the occurrence of an RNase E entry site in the *epd* ORF are most likely strongly increased by the low translation efficiency

of the *epd* ORF. Translation efficiency in bacteria depends on two main parameters: codon usage and efficiency of the TIR (49). Whereas the *pgk* and *fbaA* ORFs only contain codons corresponding to highly abundant tRNAs in *E. coli* (40), the *epd* ORF contains 9 codons corresponding to minor tRNAs, that are spread in the ORF, and one arginine codon (AGG) corresponding to a very rare tRNA in *E. coli* which is located immediately upstream of the TAA termination codon. This difference of codon usage may participate to the strong differential expression of the 3 ORFs, especially as the location of the AGG codon is a very unfavourable position, that was recently found to lead to mRNA cleavage by a yet unidentified mechanism (73). In addition, the results that we obtained with *epd-lacZ* translational and transcriptional fusions, strongly suggest that the short length between the RBS and AUG initiation codon may also participate to the low E4PDH production. Accordingly, increasing this length from 4 to 8 nts in the *epd-pgk* context increased E4PDH production by a factor of 2, without augmentation of the PGK production. A quite higher increase of the β -galactosidase activity was found when the same mutation was made in the *epd-lacZ* translational fusion of monolysogenic cells (factor of 10). In this case, the increase of translational efficiency was not counteracted by RNase E cleavage in the ORF. Taken together, these data reinforces the idea of a synergistic effect between the low efficiency of the TIR and the presence of RNase E cleavage sites in the ORF. However, the short distance between the RBS and AUG codon is probably not the only parameter responsible for the low translation efficiency of the *E. coli epd* ORF, since, in monolysogenic cells with an optimized length, the β -galactosidase activity found for the translational fusion was still significantly lower compared to that found for the transcriptional fusion (factor of 10). Also, as mentioned in the Results section, the stronger difference of β -galactosidase activity observed between transcriptional and translational fusions, in monolysogenic cells compared to transformed cells (factor of 150, against a factor of 10, respectively), strongly suggests the occurrence of a

translational repressor that is titrated in transformed cells. Hence, we proposed that the *E. coli epd* translation is negatively controlled by a transacting factor which has to be identified. The fact that the RNase E cleavages generated by a purified degradosome fraction in a transcript corresponding to the *epd* ORF are nearly identical to the RNase E cleavages obtained *in vivo* in the presence of active ribosomes, also suggests a very low level of occupancy of the *epd* ORF by translating ribosomes. Altogether, our data are in favour of a very low translational efficiency of the *epd* ORF.

The same *epd-pgk-fbaA* ORF organization as in *E. coli* is found in other *enterobacteriaceae* species e.g. *Salmonella typhi*, *Yersinia pestis* and some of the *Vibrionaceae* species e.g. *Vibrio cholerae*, *Shewanella putrefaciens* {Bardey, 2000 #24; <http://www.tigr.org/tigr-scripts/operons/operons.cgi>). Inspection of the nucleotide sequence, that are located upstream of the initiator AUG of the *epd*, *pgk* and *fbaA* ORFs in these four species (Fig. 7B), reveals a strong conservation of the short distance (4 nts) and aligned distance (7 nts) between the RBS and AUG initiation codon of the *epd* ORFs. The longer distance found in the *E. coli pgk* and *fbaA* TIRs (8 and 10-11 nts, respectively) are also highly conserved (Fig. 7B). Accordingly, the data that we obtained for the transcriptional and translational fusions built with the *E. coli lacZ* and *V. cholerae epd* genes suggest that like the *E. coli epd* TIR, the *V. cholerae epd* TIR has a low efficiency, and this may be a common characteristic feature of the *epd* ORFs found upstream of the *pgk* ORFs in γ -proteobacteria.

On an evolutionary point of view, our data suggest that the strong promoter usually present upstream of *gapA* ORF has been preserved, upon *gapA* to *epd* conversion. This may be explained by its high importance for the expression of the downstream *pgk* and *fbaA* ORFs. Indeed, promoter *gapB* P0 likely modulates the metabolic flux during both glycolysis and neoglucogenesis, especially through its regulation by the Cra protein (64, 67). As the *epd* ORF, which has not to be expressed in high yields, was under the control of promoter *gapB*

P0, the *epd* codon usage and TIR have been adapted in order to limit translation efficiency and an internal RNase E site may have been selected inside the ORF.

MATERIALS AND METHODS

Bacterial strains, media, and growth conditions.

The phenotypes of the various strains used in this study are given in Table 1. The *E. coli* TG1 strain was grown aerobically at 37°C in minimal M63 medium supplemented with 10 µg/ml thiamine, 0.1% (wt/vol) Casamino Acids and 0.3 mM ZnCl₂ (required for activity of the overproduced class II FBA (9)), and either 0.4% (wt/vol) glucose or 0.4% (wt/vol) pyruvate. The two isogenic strains used for RNase E assays, MG1693 (WT) and SK5665 (RNase E^{ts}), were grown at 30°C in Luria Broth (LB) in the presence of 50 µg/ml thymine. When the A₆₀₀ reached a value of 0.5, half of the culture was incubated at 42°C, while the other half was kept at 30°C. Samples of each culture were collected after a 15 min incubation and total RNA was prepared. The *E. coli* IBPC5321 and IBPC6311 strains, used for λ phage DNA recombination and production, respectively, were grown in LB medium supplemented with 0.01 M MgSO₄, and 0.2% maltose. Ampicillin (100 µg/ml) was added for growth of all the *E. coli* transformed cells. Growth was monitored by measuring the absorbance at 600 nm (A₆₀₀).

Construction of plasmids for the study of the *epd*, *pgk* and *fbaA* ORFs expression

Standard methods of molecular biology were used for plasmid constructions (69). The pBluescript derivatives, pPBK500, pPBK100 and pBK200, that contain the *epd* and *pgk* ORFs, with or without the *gapB* P0 promoter (Fig. 1B) and phage M13mp9:PBK100 were previously described (19). All site-directed mutageneses were performed on this phage, using the method of Kramer (42) and appropriate oligonucleotides (Table 2).

For construction of plasmids containing the three *epd*, *pgk* and *fbaA* ORFs, the genomic DNA of *E. coli* TG1 cells was extracted, and a 1.981-kb long DNA fragment, containing the 3' part of the *pgk*, the *fbaA*, and the 5' part of the *mscS* (mechanosensitive channels MscS) ORFs, was amplified by PCR with the O-GB7 and O-MSCS primers (Fig. 1A, Table 2). The amplified fragment digested with *Bam*HI, was cloned between the *Eco*RV and *Bam*HI sites of

plasmids pPBK100, and pBK200 to generate plasmids pEPF and p Δ P0-EPF, respectively. After mutation of the *pgk* P1 promoter, the 2.7-kb *Hind*III-*Bam*HI fragments of the mutated phages were cloned into plasmid pBluescript (SK)⁺ (Stratagene) to create plasmids pmut-10-PBK100 and pmutTG-PBK100. For simultaneous disruptions of the *gapB* P0 and *pgk* P1 promoters, plasmid p Δ P0-mut-10-EPF was obtained by exchange of DNA fragments between the p Δ P0-EPF and pmut-10-PBK100 plasmids.

Construction of transcriptional and translational fusions

Plasmids pRS415 and pRS414, that contain the *lacZ* ORF with or without the *lacZ* TIR, were used as recipient plasmids (71) (Fig 7C). To prepare an *E. coli* DNA fragment containing the *epd* TIR, a *Bam*HI restriction site was created in phage M13mp9:PBK100 (positions +182 and +188 in the *epd* ORF) (Fig. 1B). The two substitutions generated in the *E. coli epd* TIR (substitution of the WT sequence from +57 to +86 for the GTCGAC sequence, variant 35 and substitution of the WT sequence from +130 to +132 for the AAATAAA sequence, variant 408), were performed in phage M13mp19:PBK100mut*Bam*HI. The *Hind*III-*Bam*HI DNA fragment from WT or mutated phages M13mp19:PBK100mut*Bam*HI were inserted together with the *Mfe*I-*Hind*III fragment from plasmid pPBK500, in the recipient plasmids. The resulting pRS415 plasmid derivatives containing the WT, mut35 or mut408 transcriptional fusions were denoted pVBOF2, pVBOFv35 and pVBOFv408, respectively. Similarly, the resulting pRS414 plasmid derivatives, containing the WT and mutated translational fusions were denoted pVBGF2, pVBGF2v35 and pVBGF2v408, respectively. To create plasmids pPBK535 and pPBK5408, the 2.7 kb *Hind*III-*Bam*HI fragments of the mutated phage DNAs, were also inserted between the *Hind*III and *Bam*HI sites of plasmid pPBK500.

To built transcriptional and translational *lacZ* fusions, with the *V. cholerae epd* or *pgk* genes, two DNA fragments from the O1 N205 *V. cholerae* strain, that was isolated from an indian

choleric patient (a generous gift from P. Berche) (8), were amplified by using the oligonucleotide pair 1238-1505 or 1238-1239 (Table 2). The former fragment (*irgA*-162-*epd*+93) was used to build the *epd-lacZ* transcriptional and translational fusions (plasmids p4Vc1505 and p5Vc1505, respectively). The second fragment (*irgA*-162-*pgk*+1249) was used to build the *epd-pgk* transcriptional and translational fusions (plasmids p5Vc27 and p4Vc27, respectively). The translational and transcriptional fusions obtained with the WT or mutated *epd* genes were transferred in phage λ RS45, by homologous recombination in *E. coli* IBPC5321 cells, as previously described (71). Monolysogenic *E. coli* TG1 cells were screened by the PCR method (62). The second fragment (*irgA*-162-*pgk*+1249) was inserted at the EcoRV site of plasmid pMosBlue (Amersham Biosciences). A detailed description of all the constructs made in this study can be obtained upon request.

Plasmid copy number estimation

The plasmid copy number of the cells was estimated by extraction of the DNA contained in a defined number of cells (about 1.5×10^9 cells, corresponding to 15 A_{600}), followed by agarose gel electrophoresis and quantification of the ethidium bromid fluorescence by a Gel-Doc1000 Biorad apparatus, using the Molecular Analysis software (Biorad). The copy number of plasmid pEPF was taken as the reference.

RNA isolation

Total RNA was extracted from 30 ml of a cell culture harvested at an A_{600} of 0.5 to 0.55, by two hot phenol extractions (1), followed by two chloroform extractions and one ethanol precipitation. The RNA pellet was washed in 70% ethanol and dissolved in 70 μ l of sterile MilliQ water. The RNA concentration was estimated by A_{260} measurement before freezing at -80°C .

Northern-blot analysis

On the basis of the estimated cellular plasmid copy number, 5 to 9 μg of total RNA extracted from transformed *E. coli* TG1 cells were fractionated on a 0.8% (wt/vol) agarose-formaldehyde gel (69). Either the 16S and 23S rRNAs from the total RNA mixture or 1 μg of 5'-end labeled 100 bp DNA Ladder plus mixture (MBI Fermentas) (about 3×10^5 cpm) were used as molecular markers. The fractionated RNAs were transferred on a Biodyne B transfer membrane (0.45 μm , Pall) in 20X SSC (3 M NaCl, 3 M sodium acetate, [pH 7.0]), and were fixed by UV cross-linking for 5 min. Pre-hybridization and hybridization with the 5'-end labeled oligonucleotide O-EPD, O-PGK or O-FBA (Fig. 1A and Table 2) were performed at 42°C in 6X SSC, 0.1% SDS, 5X Denhardt's reagent, 5 mM EDTA, 10 $\mu\text{g}/\text{ml}$ sonicated and denatured calf thymus DNA (69). Hybridizations were performed with 100 ng of 5'-end labelled oligonucleotide in 10 ml of hybridization buffer. For probe labeling, 100 ng of oligonucleotide were incubated with [γ^{32} -P]ATP and T4 polynucleotide kinase, according to (19). After hybridization and before autoradiography, filters were washed in 6X SSC buffer containing 0.1% SDS, once for 2 min at room temperature, then twice for 15 min at 42°C, and once again for 2 min at room temperature.

Primer-extension analysis

Primer-extension analyzes were performed with 4 ng of 5'-end labeled oligonucleotide and 5 μg of total RNAs extracted from transformed cells or 20 μg of total RNAs extracted from untransformed cells, using conditions previously described (19). The resulting cDNAs were fractionated on a sequencing gel with the products of dideoxy-sequencing reactions obtained with the same primer.

S1 mapping experiments

The 5' end-labelled probe (Fig. 1B) used for analysis of *pgk* mRNA extremities, was amplified by PCR, by using primers O-GB1 and O-PGK, and purified on a 0.8% (wt/vol) agarose gel. About 1.5×10^5 cpm of the probe were annealed with 5 μ g of total RNA extracted from TG1 cells transformed with plasmid pPBK100 or 30 μ g of total RNA extracted from untransformed cells. The S1 nuclease treatment was performed according to Sambrook (69). The resistant DNA fragments were fractionated on a sequencing gel and compared with the products of a sequencing reaction.

***In vitro* maturation by RNase E**

A DNA fragment containing a T7 promoter was PCR amplified by using primers O-GB9 and O-GB10 (Table 2). It was then used as the matrix for *in vitro* transcription of RNA $F_{epd-pgk}$ by T7 RNA polymerase using conditions previously described (57). The $F_{epd-pgk}$ transcript was purified by filtration on a sephadex G25 column. We used a degradosome fraction from the *E. coli* SK5665 (RNase E^{ts}) strain generously provided by A. Carpousis and prepared according to (17). After a 1/32 (vol/vol) dilution in 500 mM NaCl, 5% glycerol, 0.5% Triton X-100, 1 mM EDTA, 1 mM DTT, 10 mM Tris-HCl (pH 7.5) buffer, degradosome fractions were incubated for 10 min at 30°C or 45°C (for thermal inactivation) and 2 μ l of the incubated preparations were used to digest 0.8 pmol of transcript in the conditions previously described (32). The digestion products were analyzed by extension of oligonucleotide O-PGK with reverse transcriptase (RT).

mRNA stability analysis

E. coli MG1693 (WT) cells transformed with plasmid pEPF were grown at 30°C in LB medium to an A_{600} of 0.5. Rifampicin was then added at a 0.5 mg/ml final concentration to a 150 ml fraction of the culture, 12 ml aliquots of the culture were collected before rifampicin

addition, and at different times after the addition (0, 1, 2, 4, 7, 10, 15, 20 and 30 min) (Time 0 represents the time at which rifampicin reaches RNA polymerase and inhibits transcription initiation, about 30s after rifampicin addition to the cell culture). An amount of 15 µg of total RNA extracted from each fraction was subjected to Northern–blot analysis.

RNA half-life calculations

Radioactivity in the Northern blot signals was measured with a Molecular dynamics phosphorimager (Amersham Biosciences) by using the ImageQuant software (version 5.2). This allowed us to estimate the residual amounts of *pgk-fbaA*, *pgk* and *fbaA* RNAs in each aliquots. These amounts were plotted against the times of aliquot recovery. The slopes of the straight obtained, gave an estimation of the apparent half-lives of the 3 RNAs. As a very low amount of tricistronic *epd-pgk-fbaA* RNA was present at the time of rifampicin addition, we considered that the measured half-life for the *pgk-fbaA* RNA was a real half-life. Since *fbaA* RNA only resulted from cleavage of the *pgk-fbaA* RNA, we calculated the real half-life of the *fbaA* RNA by using the two following equations:

$$d[pgk-fbaA] / dt = k_d_{pgk-fbaA} [pgk-fbaA] \quad (1)$$

$$d[fbaA] / dt = k_d_{pgk-fbaA} [pgk-fbaA] - k_d_{fbaA} [fbaA] \quad (2)$$

where $[pgk-fbaA]$ is the concentration of the bicistronic *pgk-fbaA* transcript, $[fbaA]$ is the concentrations of the monocistronic *fbaA* transcripts, $k_d_{pgk-fbaA}$ is the decay constant for the bicistronic transcript, k_d_{fbaA} is the decay constant for the monocistronic *fbaA* transcripts.

Enzymatic specific activity measurements

For measurements of the PGK and FBA activities, 30 ml fractions of cell cultures, that reached an A_{600} of 0.5 (about 1.5×10^9 cells) were centrifuged. The cell pellet was washed and sonicated in 3 ml of 50 mM Tris-HCl (pH 8.0) buffer, as previously described (19). The PGK

and FBA activities were measured by their coupling with the *Bacillus stearothermophilus* GAPDH activity. For PGK activity measurement, NADH disappearance was followed as previously described (19). The specific activity was expressed in $\mu\text{mol NADH}/\text{min}/A_{280}/A_{600}$. For class II FBA specific activity determination, we used a protocol adapted from the one of Christian *et al.* (20): 500 μl of a 3 mM fructose-1,6-bisphosphate (Sigma), 3 mM NAD^+ , 17 mM Na arsenate, 50 mM Tris-HCl (pH 8.0) solution containing 0.2 A_{280}/ml of *B. stearothermophilus* GAPDH) was preincubated 5 min at 30°C and an aliquot of the cell sonication soluble fraction (2 to 20 μl) was added. The NADH appearance was followed at 340 nm. The FBA activity was expressed in $\mu\text{mol NADH}/\text{min}/A_{280}/A_{600}$. To compare the expressions of the *pgk* and *fbaA* ORFs from plasmid pEPF, p ΔP0 -EPF and p ΔP0 -mut-10-EPF, the PGK and FBA activities measured for cells transformed with these two latter plasmids were multiplied by the ratio of the plasmid copy number in pEPF transformed cells versus that in p ΔP0 -EPF and p ΔP0 -mut-10-EPF transformed cells, respectively.

E4PDH activity was measured in the stationary phase of growth on a volume of cell culture corresponding to 60 A_{600} . The cell pellet was washed and sonicated in 3 ml of 2 mM EDTA, 50 mM Tris-HCl (pH 8.0) buffer and E4PDH specific activity was measured as described (11). The PGK specific activity was determined on the same samples, using the same conditions as above.

For measurement of the β -galactosidase-specific activity, the *E. coli* TG1 cells, transformed with the different *epd-lacZ* plasmid constructs, or lysogenic TG1 cells were harvested at A_{600} of both 0.2 and 0.4 Units. The β -galactosidase activity was measured on toluene-lysed cells, as previously described (55).

ACKNOWLEDGMENTS M. Graffe is thanked for her help in the construction of *epd-lacZ* fusions. A. J. Carpousis is greatly acknowledged for providing us with the degradosome fraction, containing the RNase E. S. Boshi-Muller and S. Azza are thanked for their gift of

Bacillus stearothermophilus GAPDH. P. Berche is thanked for his generous gift of *V. cholerae* DNA from the O1 N205 *V. cholerae* strain, isolated from an indian choleric patient. This work was supported by the French Ministère délégué à la recherche et aux nouvelles technologies, the Centre National de la Recherche Scientifique, The PRST Bioingénierie of Région Lorraine. Vincent Bardey and Corinne Vallet had a CNRS-BDI fellowship, co-financed by the Région Lorraine for Vincent Bardey. Benoit Thouvenot was supported by a fellowship from the French Ministère délégué à la recherche et aux nouvelles technologies.

REFERENCES

1. **Aiba, H., S. Adhya, and B. de Crombrughe.** 1981. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* **256**:11905-10.
2. **Alefounder, P. R., and R. N. Perham.** 1989. Identification, molecular cloning and sequence analysis of a gene cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde 3-phosphate dehydrogenase of *Escherichia coli*. *Mol. Microbiol.* **3**:723-32.
3. **Arraiano, C. M., S. D. Yancey, and S. R. Kushner.** 1988. Stabilization of discrete mRNA breakdown products in *ams pnp rnb* multiple mutants of *Escherichia coli* K-12. *J. Bacteriol.* **170**:4625-33.
4. **Bachellier, S., J. M. Clement, and M. Hofnung.** 1999. Short palindromic repetitive DNA elements in *enterobacteria*: a survey. *Res. Microbiol.* **150**:627-39.
5. **Baker, K. E., and G. A. Mackie.** 2003. Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*. *Mol. Microbiol.* **47**:75-88.
6. **Bardey, V.** 2000. Organisation et régulation de l'expression des gènes de la Glyceraldéhyde-3-phosphate déshydrogénase, la Phosphoglycérate kinase et l'Erythrose-4-phosphate déshydrogénase chez les protéobactéries de la subdivision gamma. Université Henri Poincaré, Nancy I, Nancy.
7. **Barne, K. A., J. A. Bown, S. J. Busby, and S. D. Minchin.** 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J.* **16**:4034-40.
8. **Berche, P., C. Poyart, E. Abachin, H. Lelievre, J. Vandepitte, A. Dodin, and J. M. Fournier.** 1994. The novel epidemic strain O139 is closely related to the pandemic strain O1 of *Vibrio cholerae*. *J. Infect. Dis.* **170**:701-4.
9. **Berry, A., and K. E. Marshall.** 1993. Identification of zinc-binding ligands in the class II fructose-1,6-bisphosphate aldolase of *Escherichia coli*. *FEBS Lett.* **318**:11-6.
10. **Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-74.
11. **Boschi-Muller, S., S. Azza, D. Pollastro, C. Corbier, and G. Branlant.** 1997. Comparative enzymatic properties of GapB-encoded erythrose-4-phosphate dehydrogenase of *Escherichia coli* and phosphorylating glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* **272**:15106-12.
12. **Bown, J. A., K. A. Barne, S. D. Minchin, and S. J. W. Busby.** 1997. Extended -10 promoters. *In* F. Eckstein and D. M. J. Lilley (ed.), *Nucleic acids and molecular biology*. Springer, Berlin.
13. **Branlant, C., T. Oster, and G. Branlant.** 1989. Nucleotide sequence determination of the DNA region coding for *Bacillus stearothermophilus* glyceraldehyde-3-phosphate dehydrogenase and of the flanking DNA regions required for its expression in *Escherichia coli*. *Gene* **75**:145-55.
14. **Branlant, G., and C. Branlant.** 1985. Nucleotide sequence of the *Escherichia coli* *gap* gene. Different evolutionary behavior of the NAD⁺-binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* **150**:61-6.
15. **Branny, P., F. de la Torre, and J. R. Garel.** 1998. An operon encoding three glycolytic enzymes in *Lactobacillus delbrueckii* subsp. *bulgaricus*: glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase. *Microbiology* **144** (Pt 4):905-14.

16. **Burchhardt, G., K. F. Keshav, L. Yomano, and L. O. Ingram.** 1993. Mutational analysis of segmental stabilization of transcripts from the *Zymomonas mobilis gap-pgk* operon. *J. Bacteriol.* **175**:2327-33.
17. **Carpousis, A. J., G. Van Houwe, C. Ehretsmann, and H. M. Krisch.** 1994. Copurification of *E. coli* RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell* **76**:889-900.
18. **Carroll, P. A., G. Zhao, S. A. Boyko, M. E. Winkler, and S. B. Calderwood.** 1997. Identification, sequencing, and enzymatic activity of the erythrose-4-phosphate dehydrogenase gene of *Vibrio cholerae*. *J. Bacteriol.* **179**:293-6.
19. **Charpentier, B., V. Bardey, N. Robas, and C. Branlant.** 1998. The EIIGlc protein is involved in glucose-mediated activation of *Escherichia coli gapA* and *gapB-pgk* transcription. *J. Bacteriol.* **180**:6476-83.
20. **Christian, W.** 1955. Aldolase from yeast, p. 315-320, *Methods Enzymol.*, vol. 1.
21. **Condon, C., C. Squires, and C. L. Squires.** 1995. Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* **59**:623-45.
22. **Conway, T., and L. O. Ingram.** 1988. Phosphoglycerate kinase gene from *Zymomonas mobilis*: cloning, sequencing, and localization within the *gap* operon. *J. Bacteriol.* **170**:1926-33.
23. **Della Seta, F., S. Boschi-Muller, M. L. Vignais, and G. Branlant.** 1997. Characterization of *Escherichia coli* strains with *gapA* and *gapB* genes deleted. *J. Bacteriol.* **179**:5218-21.
24. **Ehretsmann, C. P., A. J. Carpousis, and H. M. Krisch.** 1992. Specificity of *Escherichia coli* endoribonuclease RNase E: in vivo and in vitro analysis of mutants in a bacteriophage T4 mRNA processing site. *Genes Dev.* **6**:149-59.
25. **Eikmanns, B. J.** 1992. Identification, sequence analysis, and expression of a *Corynebacterium glutamicum* gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomerase. *J. Bacteriol.* **174**:6076-86.
26. **Espeli, O., L. Moulin, and F. Bocard.** 2001. Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J. Mol. Biol.* **314**:375-86.
27. **Fillinger, S., S. Boschi-Muller, S. Azza, E. Dervyn, G. Branlant, and S. Aymerich.** 2000. Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. *J. Biol. Chem.* **275**:14031-7.
28. **Fraenkel, D. G.** 1996. glycolysis, p. 189-198. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium* : cellular and molecular biology 2nd edn., vol. 1, Washington D.C.
29. **Gebbia, J. A., P. B. Backenson, J. L. Coleman, P. Anda, and J. L. Benach.** 1997. Glycolytic enzyme operon of *Borrelia burgdorferi*: characterization and evolutionary implications. *Gene* **188**:221-8.
30. **Grunberg-Manago, M.** 1999. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu. Rev. Genet.* **33**:193-227.
31. **Gschaedler, A., N. Robas, J. Boudrant, and C. Branlant.** 1999. Effects of pulse addition of carbon sources on continuous cultivation of *Escherichia coli* containing a recombinant *E. coli gapA* gene. *Biotechnol. Bioeng.* **63**:712-20.
32. **Hajnsdorf, E., O. Steier, L. Coscoy, L. Teyssset, and P. Regnier.** 1994. Roles of RNase E, RNase II and PNPase in the degradation of the *rpsO* transcripts of *Escherichia coli*: stabilizing function of RNase II and evidence for efficient degradation in an *ams pnp rnb* mutant. *EMBO J.* **13**:3368-77.
33. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* **15**:2343-61.

34. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477-83.
35. **Hillman, J. D., and D. G. Fraenkel.** 1975. Glyceraldehyde 3-phosphate dehydrogenase mutants of *Escherichia coli*. *J. Bacteriol.* **122**:1175-9.
36. **Iost, I., and M. Dreyfus.** 1995. The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.* **14**:3252-61.
37. **Jiang, X., and J. G. Belasco.** 2004. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proc. Natl. Acad. Sci. U S A* **101**:9211-6.
38. **Kaberdin, V. R.** 2003. Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res.* **31**:4710-6.
39. **Kaberdin, V. R., A. P. Walsh, T. Jakobsen, K. J. McDowall, and A. von Gabain.** 2000. Enhanced cleavage of RNA mediated by an interaction between substrates and the arginine-rich domain of *E. coli* ribonuclease E. *J. Mol. Biol.* **301**:257-64.
40. **Karlin, S., J. Mrazek, A. Campbell, and D. Kaiser.** 2001. Characterizations of highly expressed genes of four fast-growing bacteria. *J. Bacteriol.* **183**:5025-40.
41. **Keilty, S., and M. Rosenberg.** 1987. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J. Biol. Chem.* **262**:6389-95.
42. **Kramer, W., V. Drutsa, H. W. Jansen, B. Kramer, M. Pflugfelder, and H. J. Fritz.** 1984. The gapped duplex DNA approach to oligonucleotide-directed mutation construction. *Nucleic Acids Res.* **12**:9441-56.
43. **Kumar, A., R. A. Malloch, N. Fujita, D. A. Smillie, A. Ishihama, and R. S. Hayward.** 1993. The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J. Mol. Biol.* **232**:406-18.
44. **Lee, K., J. A. Bernstein, and S. N. Cohen.** 2002. RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. *Mol. Microbiol.* **43**:1445-56.
45. **Lesnik, E. A., R. Sampath, H. B. Levene, T. J. Henderson, J. A. McNeil, and D. J. Ecker.** 2001. Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res.* **29**:3583-94.
46. **Li, Y., and S. Altman.** 2003. A specific endoribonuclease, RNase P, affects gene expression of polycistronic operon mRNAs. *Proc. Natl. Acad. Sci. U S A* **100**:13213-8.
47. **Lisser, S., and H. Margalit.** 1993. Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res.* **21**:1507-16.
48. **Ludwig, H., G. Homuth, M. Schmalisch, F. M. Dyka, M. Hecker, and J. Stulke.** 2001. Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the *gapA* operon. *Mol. Microbiol.* **41**:409-22.
49. **Ma, J., A. Campbell, and S. Karlin.** 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**:5733-45.
50. **Mackie, G. A.** 1998. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**:720-3.

51. **Marchand, I., A. W. Nicholson, and M. Dreyfus.** 2001. Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. *Mol. Microbiol.* **42**:767-76.
52. **Martin, W., H. Brinkmann, C. Savonna, and R. Cerff.** 1993. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl. Acad. Sci. U S A* **90**:8692-6.
53. **Meijer, W. G., E. R. van den Bergh, and L. M. Smith.** 1996. Induction of the *gap-pgk* operon encoding glyceraldehyde-3-phosphate dehydrogenase and 3-phosphoglycerate kinase of *Xanthobacter flavus* requires the LysR-type transcriptional activator CbbR. *J. Bacteriol.* **178**:881-7.
54. **Meinken, C., H. M. Blencke, H. Ludwig, and J. Stulke.** 2003. Expression of the glycolytic *gapA* operon in *Bacillus subtilis*: differential syntheses of proteins encoded by the operon. *Microbiology* **149**:751-61.
55. **Miller, J. H.** 1972. *Experiments in Molecular Genetics.*, Cold Spring Harbor, N.Y.
56. **Mitchell, J. E., D. Zheng, S. J. Busby, and S. D. Minchin.** 2003. Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res.* **31**:4689-95.
57. **Mougin, A., A. Gregoire, J. Banroques, V. Segault, R. Fournier, F. Brule, M. Chevrier-Miller, and C. Branlant.** 1996. Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA* **2**:1079-93.
58. **Nellemann, L. J., F. Holm, T. Atlung, and F. G. Hansen.** 1989. Cloning and characterization of the *Escherichia coli* phosphoglycerate kinase (*pgk*) gene. *Gene* **77**:185-91.
59. **Peng, L., and K. Shimizu.** 2003. Global metabolic regulation analysis for *Escherichia coli* K12 based on protein expression by 2-dimensional electrophoresis and enzyme activity measurement. *Appl. Microbiol. Biotechnol.* **61**:163-78.
60. **Plumbridge, J., and D. Soll.** 1989. Characterization of cis-acting mutations which increase expression of a *glnS-lacZ* fusion in *Escherichia coli*. *Mol. Gen. Genet.* **216**:113-9.
61. **Plumbridge, J. A., J. Dondon, Y. Nakamura, and M. Grunberg-Manago.** 1985. Effect of NusA protein on expression of the *nusA, infB* operon in *E. coli*. *Nucleic Acids Res.* **13**:3371-88.
62. **Powell, B. S., M. P. Rivas, D. L. Court, Y. Nakamura, and C. L. Turnbough, Jr.** 1994. Rapid confirmation of single copy lambda prophage integration by PCR. *Nucleic Acids Res.* **22**:5765-6.
63. **Prost, J. F., and A. J. Cozzone.** 1999. Detection of an extended-10 element in the promoter region of the *pckA* gene encoding phosphoenolpyruvate carboxykinase in *Escherichia coli*. *Biochimie* **81**:197-200.
64. **Ramseier, T. M., S. Bledig, V. Michotey, R. Feghali, and M. H. Saier, Jr.** 1995. The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. *Mol. Microbiol.* **16**:1157-69.
65. **Redko, Y., M. R. Tock, C. J. Adams, V. R. Kaberdin, J. A. Grasby, and K. J. McDowall.** 2003. Determination of the catalytic parameters of the N-terminal half of *Escherichia coli* ribonuclease E and the identification of critical functional groups in RNA substrates. *J. Biol. Chem.* **278**:44001-8.
66. **Romby, P., and M. Springer.** 2003. Bacterial translational control at atomic resolution. *Trends Genet.* **19**:155-61.
67. **Saier, M. H., Jr., and T. M. Ramseier.** 1996. The catabolite repressor/activator (Cra) protein of enteric bacteria. *J. Bacteriol.* **178**:3411-7.

68. **Salgado, H., S. Gama-Castro, A. Martinez-Antonio, E. Diaz-Peredo, F. Sanchez-Solano, M. Peralta-Gil, D. Garcia-Alonso, V. Jimenez-Jacinto, A. Santos-Zavaleta, C. Bonavides-Martinez, and J. Collado-Vides.** 2004. RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res.* **32 Database issue:**D303-6.
69. **Sambrook, J., F. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
70. **Schlapfer, B. S., and H. Zuber.** 1992. Cloning and sequencing of the genes encoding glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase (*gap* operon) from mesophilic *Bacillus megaterium*: comparison with corresponding sequences from thermophilic *Bacillus stearothermophilus*. *Gene* **122:**53-62.
71. **Simons, R. W., F. Houman, and N. Kleckner.** 1987. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* **53:**85-96.
72. **Stokes, N. R., H. D. Murray, C. Subramaniam, R. L. Gourse, P. Louis, W. Bartlett, S. Miller, and I. R. Booth.** 2003. A role for mechanosensitive channels in survival of stationary phase: regulation of channel expression by RpoS. *Proc. Natl. Acad. Sci. U S A* **100:**15959-64.
73. **Sunohara, T., K. Jojima, Y. Yamamoto, T. Inada, and H. Aiba.** 2004. Nascent-peptide-mediated ribosome stalling at a stop codon induces mRNA cleavage resulting in nonstop mRNA that is recognized by tmRNA. *RNA* **10:**378-86.
74. **Thouvenot, B., B. Charpentier, and C. Branlant.** 2004. The strong efficiency of the *Escherichia coli gapA* P1 promoter depends upon a complex combination of functional determinants. *Biochem. J.* **Pt.**
75. **Wachi, M., and K. Nagai.** 2001. *Escherichia coli* ribonuclease G., p. 55-63, *Methods Enzymol.*, vol. 342.
76. **Wosten, M. M.** 1998. Eubacterial sigma-factors. *FEMS Microbiol. Rev.* **22:**127-50.
77. **Yang, Y., G. Zhao, T. K. Man, and M. E. Winkler.** 1998. Involvement of the *gapA*- and *epd* (*gapB*)-encoded dehydrogenases in pyridoxal 5'-phosphate coenzyme biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.* **180:**4294-9.
78. **Zhao, G., A. J. Pease, N. Bharani, and M. E. Winkler.** 1995. Biochemical characterization of *gapB*-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and its possible role in pyridoxal 5'-phosphate biosynthesis. *J. Bacteriol.* **177:**2804-12.

FIGURE LEGENDS

Figure 1: Schematic representation of the *epd-pgk-fbaA* ORF cluster and of the different constructs used in this study

Panel A: the *epd-pgk-fbaA* ORF cluster, bordered by the *yggC* (putative kinase) and *mscS* (mechanosensitive channel MscS) ORFs, is depicted. White rectangles represent untranslated regions. The ORFs are shown by gray or black rectangles. The *gapB* P0 (19), *pgk* P1 (this study) and *mscS* P (72) transcription-start sites are marked by bent arrows. The potential terminators are symbolized by thin lines surmounted of gray circles. The target sequences of the oligonucleotides used in this study are represented by black triangles, their orientations indicate the complementary strand (▶non coding, ◀coding strand). Numbering is according to the complete sequence of the *E. coli* K12 chromosome (10). Positions of cleavage by restriction enzymes are indicated in parentheses.

Panel B: The amplified DNA fragments used for production of the $F_{epd-pgk}$ and $F_{pgk-fbaA}$ RNA fragments by *in vitro* transcription are shown. The horizontal black thin line represents the 5' end-labeled probe used for the S1-nuclease mapping experiments. The star indicates the labeled extremity. The regions of the *epd-pgk-fbaA* ORF cluster present in plasmids pEPF, p Δ P0-EPF, p Δ P0-mut-10P1-EPF, pPBK500, pPBK100 and pBK200 are represented with the same rules of colors as in panel A. The mutations generated in plasmid pPBK500 (mut35 and mut408 in the *epd* TIR), and in pPBK100 (mutTG and mut-10 in promoter *pgk* P1) are indicated. In plasmid p Δ P0-mut-10P1-EPF, the X symbol indicates the mut-10 mutation present in promoter *pgk* P1. In panel B, the *gapB* P0 transcription start site is taken as the reference for bp numbering: bp or nts located downstream from the start site are numbered from +2 to +4089, those located upstream are number from -1 to -680.

Figure 2: Northern-blot analysis of the *epd-pgk-fbaA*, *epd-pgk*, *pgk-fbaA* transcripts in *E. coli* TG1 cells transformed with plasmids pEPF, p Δ P0-EPF, p Δ P0-mut-10P1-EPF, or pBluescript (pBS)

The transformed cells were grown in M63 medium in the presence of glucose until an A_{600} of 0.5. For each transformed strain, total RNA and DNA were extracted from two distinct 30 ml fractions of the cell culture. The amount of total RNA, that was loaded on the 0.8% agarose-formaldehyde gel was adjusted taking into account the plasmid copy number of the transformed cells (8.6, 4.3, 6.4 and 6 μ g were loaded for cells transformed with plasmids pEPF, p Δ P0-EPF, p Δ P0-mut-10P1-EPF, and pBS, respectively). After electrophoresis, the RNAs were transferred on a biodyne membrane. Successive hybridizations of the membrane were performed with the 5'-end labeled oligonucleotides O-EPD (panel A), O-PGK (panel B), and O-FBA (panel C). The names of the plasmids used are indicated at the top of the lanes. The identity of the hybridization products was deduced from their estimated lengths and the probes to which they hybridize. The experimental lengths were estimated, by using the *E. coli* 16S and 23S rRNAs as molecular markers. The identities of the hybridization products, that were deduced by comparison with the predicted lengths given in panel D, are given on the left side of the autoradiograms.

Panel D: Schematic representation of the possible transcriptional organization of the *epd-pgk-fbaA* ORF cluster and of possible transcription and maturation products with their predicted lengths in kbs. The *epd-pgk-fbaA* ORF cluster, its transcription start-site and putative transcription terminators are represented as in Figure 1A.

Figure 3: Characterization of the 5' extremities of *pgk* mRNAs

Panels A₁ and A₂: Primer-extension analyses of total RNAs (5 μ g) extracted from TG1 cells grown in M63 medium in the presence of glucose and transformed with the pBK200 (lane 1), pmutTG-PBK100 (lane 2), pmut-10-PBK100 (lane 3), or pPBK100 (lane 4) plasmids using

oligonucleotide O-PGK (panel A₁) or O-EPD (panel A₂) were used as the primers. Sequencing ladders (lanes A, C, G and U) were obtained with the same oligonucleotides. The 5' extremities of *pgk* mRNAs (*pgk* 1 to 6) and their positions as referred to the *gapB* P0 start site are indicated on the left and right sides of the figure, respectively.

Panels B₁ and B₂: S1 mapping assay on total RNA extracted from untransformed (30 µg) (lane 1) or pEPF transformed (5 µg) (lane 2) TG1 cells grown in M63 medium supplemented with glucose. The 5'-end labeled O-GB1/O-PGK DNA fragment (Fig. 1B) was used as the probe. Lanes U, G, C, and A correspond to the sequencing ladder obtained with oligonucleotide O-PGK. The designation of the identified *pgk* mRNA 5' extremities and their positions as referred to the *gapB* P0 start site are indicated on the left and right sides of the autoradiograms, respectively.

Panel C₁ and C₂: Effect of the identity of the carbon source on the *pgk* mRNA composition. Untransformed TG1 cells were grown in M63 medium supplemented with glucose (lane 1) or pyruvate (lane 2), 20 µg of the total RNAs, extracted in the mid-log phase of growth, were analyzed by extension of oligonucleotide O-PGK (panel E), or O-EPD (panel F) with RT. The *pgk* mRNA numbering is as in Panel A.

Panel D: Schematic representation of the data shown in Panels A and B on the *epd-pgk* nucleotide sequence. The 5' extremity of the *pgk* P1 mRNA is shown by a bent arrow. Those of the cleavage products are shown by triangles. The *epd* and *pgk* ORFs are represented by triplets of nucleotides. The *epd* TAA stop codon and the *pgk* ATG initiation codon are indicated in bold. The extended -10 sequence upstream of site 1 (*pgk* P1 promoter) is boxed. The sequence complementary to oligonucleotide O-PGK is underlined.

Figure 4: The *pgk-fbaA* interORF region is cleaved by RNase E

Panel A: Identification of the putative 5' extremities of *pgk* mRNAs from pEPF transformed TG1 cells by extension of the O-PGK probe with RT (lane RT). Several cleavages were

detected in the run of A residues from position +2364 to +2369 (*fbaA* 5 series of sites). Lanes U, G, C, and A correspond to the sequencing ladder. Positions of the detected RT stops as referred to the *gapB* P0 start site are given on the right side of the autoradiogram.

Panel B: RT analysis with oligonucleotide O-FBA of the total RNA (5µg) extracted from pEPF (lane 1) and pΔP0-EPF (lane 2) transformed TG1 cells, that were grown in M63 medium supplemented with glucose, and of the *in vitro* transcribed F_{*pgk-fbaA*} RNA (30 ng) (Fig. 1B). The cDNA bands are numbered as in panel A.

Panel C: MG1693 (WT) (lanes 1 and 2) and SK5665 (RNase E^{ts}) (lanes 3 and 4) *E. coli* cells transformed with plasmid pEPF were grown at 30°C in LB medium. One half of each culture was incubated for 15 min at 42°C (lanes 2 and 4), and the other half was maintained at 30°C (lanes 1 and 3). Total RNA extracted from each cells batches (5 µg) were analyzed by extension of primer O-FBA with RT. The resulting *fbaA* cDNAs are numbered as in Panel A.

Panel D: Schematic representation of the data shown in Panel A and B on the nucleotide sequence of the *pgk-fbaA* intergenic region. The series of *fbaA* 5 cleavage sites are indicated by an horizontal line surmounted of a triangle and the *fbaA* 2 cleavage site is shown by a triangle. The *pgk* and *fbaA* ORFs are represented by triplets of nucleotides and the *pgk* TAA stop codon and the *fbaA* ATG initiation codon are in bold. The two horizontal inverted arrows represent the potential *pgk* transcription terminator. The -10 and -35 hexamers of the previously proposed *fbaA* transcriptional promoter (68) are boxed in gray.

Figure 5: *In vivo* and *in vitro* cleavages of *epd-pgk* transcripts by RNase E

Panel A: *In vivo* maturation. The MG1693 (WT) and SK5665 (RNase E^{ts}) *E. coli* strains transformed with plasmid pEPF were grown at 30°C in LB medium, until an A₆₀₀ of 0.5, and one half of each culture was subjected to a temperature shift at 42°C. Total RNAs extracted from the different batches of cells were analyzed by extension of primer O-PGK with RT. Lane U, G, C and A corresponds to the sequencing ladder.

Panel B: *In vitro* maturation assay. The *in vitro* transcribed $F_{epd-pgk}$ RNA (Fig. 1) was incubated 10 min with a degradosome fraction, prepared from the *E. coli* RNase E^{ts} strain, that was preincubated either 10 min at 45°C (lane 1) or 10 min at 30°C (lane 2). The reaction products were analyzed by extension of primer O-PGK with RT. As a control, the same RT analysis was performed on total RNA (5 µg) from pPBK100 transformed cells (lane 3). Lanes U, G, C, and A correspond to the sequencing ladder. The RNase E cleavage sites obtained *in vitro* are numbered from E2 to E6.

Panel C: Schematic representation of the data shown in Panels A and B on the nucleotide sequence of the *epd-pgk* region from position +874 to +1243. The 5' extremities *pgk* mRNAs 2 to 6, corresponding to *in vivo* RNase E cleavage, are indicated by triangles drawn below the sequence. The positions of the *in vitro* maturation sites are indicated by triangles drawn above the sequence. The extended -10 box of promoter *pgk* P1 is boxed and the corresponding transcription start site is shown.

Figure 6: Individual *fbaA* mRNAs are more stable than individual *pgk* mRNAs.

Panel A: Northern-Blot analysis of total RNAs extracted from aliquots of *E. coli* TG1 cells transformed with plasmid pEPF, grown in LB, and treated with rifampicin, as described in Materials and Methods. Samples were collected 1 min before, 0, 1, 2, 4, 7, 15, 20 and 30 min after the addition (the times of sample collection are indicated above the lanes). The total RNAs extracted from the various cell samples were fractionated by gel electrophoresis and transferred on a biodyne membrane. Hybridization of the same membrane was successively done with the 5' end-labelled oligonucleotide O-EPD (panel A), O-PGK (panel B) and O-FBA (panel C).

Panel B: Quantification of the data shown in Panel A was achieved by measuring the radioactivity in the bands with a phosphorimager. The amount of residual *pgk-fbaA* (Subpanel B₁ and B₂), *pgk* (Subpanel B₁) and *fbaA* (Subpanel B₂) RNAs expressed as a percentage of

their value at the time of rifampicin addition were plotted against the times of sample collection.

Figure 7: The strategy used to study the *E. coli* and *V. cholerae epd* TIRs

Panel A: The *E. coli* chromosomal DNA region used for building the *epd-lacZ* transcriptional and translational fusions is shown, using the same symbols as in figure 1. The TIR is shown by a grey ovoid in a black rectangle. The *Bam*HI restriction site generated between position +182 and +188 is indicated. The *E. coli* DNA fragment (position -583 to +182) present in the pRS414 and pRS415 plasmid derivatives, is represented by a thin line.

Panel B: The *V. cholerae* chromosomal DNA region used for building the *epd-lacZ* (position +1921 from *irgA* ORF to +48 from *epd* ORF) and *epd-pgk-lacZ* (position +1921 from *irgA* ORF to +47 from *pgk* ORF) transcriptional and translational fusions are shown. The *irgA* (major iron-regulated outer membrane protein), *epd* and *pgk* ORFs are represented by black, grey and dark grey rectangles, respectively. Non coding sequences are in white and the TIRs are represented as in panel A. The two *V. cholerae* DNA fragments inserted in the pRS414 and pRS415 plasmid derivatives are represented by two thin lines.

Panel C: The regions of the pRS414 and pRS415 plasmids in which the *E. coli* or *V. cholerae* DNA fragments were inserted are shown. The *lacZ* ORF is dashed, the *lacZ* TIR is represented with the same symbol as the *epd* TIR in Panel A. The *Eco*RI and *Bam*HI sites used for the insertion are shown. The 4 *rrnB* terminators upstream of the *lacZ* ORF are indicated.

Panel D: Model of secondary structure for the 5'-terminal region of the *E. coli epd* transcript. The long stem-loop structure upstream of the RBS element is shown (2). The RBS and AUG initiation codon of the *epd* ORF are in bold. The sequence substitutions achieved in variants 35 and 408 are indicated by the boxed sequences marked by 35 and 408, respectively.

Panel E: Sequence alignment of the *V. cholerae*, *E. coli*, *S. typhi* and *Y. pestis epd* TIR regions. The aligned sequences are from the TIGR web site (<http://www.tigr.org>). Residues complementary to the anti-RBS sequence of 16 rRNA are boxed in grey. The initiator AUG is shown in dark grey. The distance and aligned distance between the RBS and initiator AUG, as defined by (49), are given on the right side of the panel, respectively.

Figure 8: Transcriptional organization and maturation processes of the *E. coli epd-pgk-fbaA* operon.

Panel A: The same representation as in Figure 1 is used for the *epd-pgk-fbaA E. coli* DNA region. The four transcripts produced from the *epd-pgk-fbaA* DNA region are shown. RNase E cleavage sites are represented by scissors and ribosomes translating the ORFs are shown. The internal entry site is indicated by a triangle drawn upstream of the transcript.

Panel B: Secondary structure of the RNA regions containing the *pgk 2* (subpanel B₁) and *fbaA 2* sites and the *fbaA 5* series of sites (subpanel B₂). These secondary structures were determined experimentally (Bardey and Vallet, unpublished results). The *pgk* P1 transcription start site and the putative rho-independent *pgk* transcription terminator are shown. The *pgk* TAA stop codon is indicated in bold (subpanel B₂).

Panel C: The stable stem-loop structure located at the 3' extremity of the 484-nt long resistant fragment corresponding to the 5' extremity of *epd* transcripts is shown.

Panel D: Predicted secondary structure of the REP element located at the 3' extremity of the *fbaA* ORF, the three putative rho-independent and the putative rho-dependent transcription terminators are shown.

TABLE 1: The *E. coli* bacterial strains used in this study

Bacterial strains	Relevant genotypes	Source or origin
TG1	<i>supE hsdΔ5 thi Δ(lac proAB)</i> , F ⁺ (<i>traD36 proAB⁺ lacIq lacZΔM15</i>)	Stratagene
IBPC5321	<i>thi-1 argE3 ΔlacX74 mlt-1 xyl-5 tsx-29 rpsL argG6 his-4</i>	(61)
IBPC6311	<i>thi-1 argE3 ΔlacX74 mlt-1 xyl-5 tsx-29 rpsL argG6 his-4</i> <i>recA56 srl-300::Tn10</i>	(60)
MG1693	<i>thyA715</i>	(3)
SK5665	<i>thyA715 ams-1</i>	(3)
IBPC602	<i>thi-1 argE3 ΔlacX74 mlt-1 xyl-5 rpsL argG6 his-4 nadA51::Tn10</i>	P. Régnier
IBPC600	<i>thi-1 argE3 ΔlacX74 mlt-1 xyl-5 rpsL argG6 his-4 nadA51::Tn10, rnc105</i>	P. Régnier
N3433	HfrH <i>lacZ43 λ::relA1 spoT thi-1</i>	D. Apirion
IBPC935	HfrH <i>lacZ43 λ::relA1 spoT thi-1 rng::CmR</i>	E. Hajsndorf

TABLE 2: Nucleotide sequence^a of the oligonucleotide primers used in this study

<i>Primers used for PCR amplification</i>		<i>E. coli</i> targeted sequences ^b
O-GB1	TGGAATAAAGCTTCCCACAA	(- 110) - (- 91)
O-GB7	ATCGGTGATGCTTCCGCT	(+2057) - (+2074)
O-GB8 ^c	<i>TCTAATACGACTCACTATAGGGAAAATCTCCTACATCTCCAC</i>	(+2264) - (+2283)
O-GB9 ^c	<i>TCTAATACGACTCACTATAGGGTTATTGATGCATACCATC</i>	(+701) - (+718)
O-GB10	CAACGCCGTCGAGGTAA	(+1499) - (+1483)
O-MSCS ^d	<i>CCCGCCGGCGGATCCGCAGCAGCGCCTGGTTA</i>	(+4089) - (+4073)
<i>Primers used for site-directed mutageneses</i>		
mut-10	TCCGTATAGTCAGCTGTGCCATGAAATGCA	(+971) - (+942)
mutTG	AGTCAACTATACGTTGAAATGCACCT	(+964) - (+939)
mutBamH1	ATACAAAGCAGGATCCACATTACG	(+198) - (+175)
mut35	TTGCCAGACAGAGTAGTCGACCTGTCGCAAACCTGCG	(+101) - (+43)
mut408	ACGGTACGGTCATTTTATTTCTCCTGCAAGGT	(+146) - (+117)
<i>Primers used for primer extension analyses and Northern-Blot analyses</i>		
O-EPD	AGCACGAACCACATTACG	(+201) - (+175)
O-PGK	TCCCAGCAAGATCCAGATCG	(+1241) - (+1222)
O-FBA	TCACCAGTGATTACGCCAG	(+2626) - (+2606)
<i>Primers used for PCR amplification</i>		<i>V. cholerae</i> targeted sequences ^b
1238	GGCCGCCGCTACTGGTTG	(-162) - (-145)
1239	GGATCCACGCGTTTACCTGCAAGA	(+93) - (+65)
1505	TCATACACTGCAGGATCCACGTTACGCCCA	(+1249) - (+1220)

^a The nucleotide sequence are written from 5' to 3'. ^b The positions of the extremities of the targeted sequences in both *E. coli* and *V. cholerae* DNAs are given as referred to the *gapB* P0 transcription start site taken as position +1 (Fig. 1B and 7B). ^cThe T7 promoter sequence is in italic. ^dThe oligonucleotide sequence flanking the segment complementary to the targeted *mscS* ORF sequence is in italic.

TABLE 3: The PGK and FBA specific activities^b of *E. coli* TG1 transformed cells^a

Tested construct	PGK ^c ($\mu\text{mol NADH}/\text{min}/ A_{280}/A_{600}$)	FBA ^c ($\mu\text{mol NADH}/\text{min}/ A_{280}/A_{600}$)
pEPF	2.5	0.031
p Δ P0-EPF	1.05 \pm 0.11	0.017 \pm 0.002
p Δ P0-mut-10P1-EPF	0.14 \pm 0.09	0.0015 \pm 0.001
pBS	0.10 \pm 0.09	0.001 \pm 0.001

^a *E. coli* TG1 cells transformed with plasmids pEPF, p Δ P0-EPF, p Δ P0-mut-10P1-EPF or pBS were grown in M63 minimal medium supplemented with glucose and Casamino Acids. ^b Specific activities were measured on cells harvested in the mid-log phase of growth. ^c The reported values are means and standard deviations of three independent experiments.

TABLE 4: The β -galactosidase specific activities produced from *E. coli epd-lacZ* fusions in *E. coli* TG1 transformed cells or monolysogenic cells^a.

Tested TIR	β -galactosidase Activity (Miller Unit per A_{600}) ^b			
	Plasmid		monolysogenic	
	translational fusion	transcriptional fusion	translational fusion	transcriptional fusion
<i>E. coli epd</i> WT	800+/-150	9150+/-2000	8+/-3	1230+/-96
<i>E. coli epd</i> 35 ^c	1400+/-400	9300+/-800	11+/-3	1100+/-66
<i>E. coli epd</i> 408 ^c	4300+/-500	8600+/-1700	110+/-17	1380+/-110

^a *E. coli* TG1 cells transformed with plasmids pRS414 (translational fusion) or pRS415 (transcriptional fusion) derivatives (Fig. 7) and monolysogenic *E. coli* TG1 cells containing unique copies of the translational or transcriptional fusions were grown in M63 minimal medium supplemented with glucose and Casa Aminoacids until mid-log phase of growth. ^b β -galactosidase activity was measured using the method of Miller. ^c The *E. coli epd35-lacZ* and *epd408-lacZ* fusions contained the two TIR mutations that we generated in the *epd* ORF, respectively (Fig. 7)

TABLE 5: E4PDH and PGK specific activities^b of *E. coli* TG1 transformed cells^a

Plasmid	identity of <i>epd</i> ORF	E4PDH ^c ($\mu\text{mol NADH}/\text{min}/A_{280}/A_{600}$)	PGK ^c ($\mu\text{mol NADH}/\text{min}/A_{280}/A_{600}$)
pPBK500	<i>E. coli epd</i> WT	0.47 \pm 0.15	203 \pm 10
pPBK535	<i>E. coli epd</i> 35	0.49 \pm 0.03	170 \pm 16
pPBK5405	<i>E. coli epd</i> 408	0.98 \pm 0.03	190 \pm 11
pVc27	<i>V. cholerae epd</i> WT	0.46 \pm 0.04	ND

^a *E. coli* TG1 cells transformed with plasmids pPBK500, pPBK535, pPBK5405 or pVc27 were grown in M63 minimal medium supplemented with glucose and Casamino Acids. ^b Specific activities were measured on cells harvested in the mid-log phase of growth. ^c The reported values are means and standard deviations of three independent experiments.

TABLE 6: β -galactosidase specific activities produced by *V. cholerae epd-lacZ* and *V. cholerae pgk-lacZ* fusions in *E. coli* TG1 transformed cells^a.

Tested Gene	β -galactosidase Activity (Miller Unit per A_{600}) ^b	
	translational fusion	transcriptional fusion
<i>V. cholerae epd</i>	30+/-8	435+/-35
<i>V. cholerae pgk</i>	686+/-85	860+/-93

^a *E. coli* TG1 cells transformed with plasmids pRS414 (translational fusion) or pRS415 (transcriptional fusion) derivatives (Fig. 7) were grown in M63 minimal medium supplemented with glucose and Casa Aminoacids until mid-log phase of growth. ^b β -galactosidase activities was measured using the method of Miller.

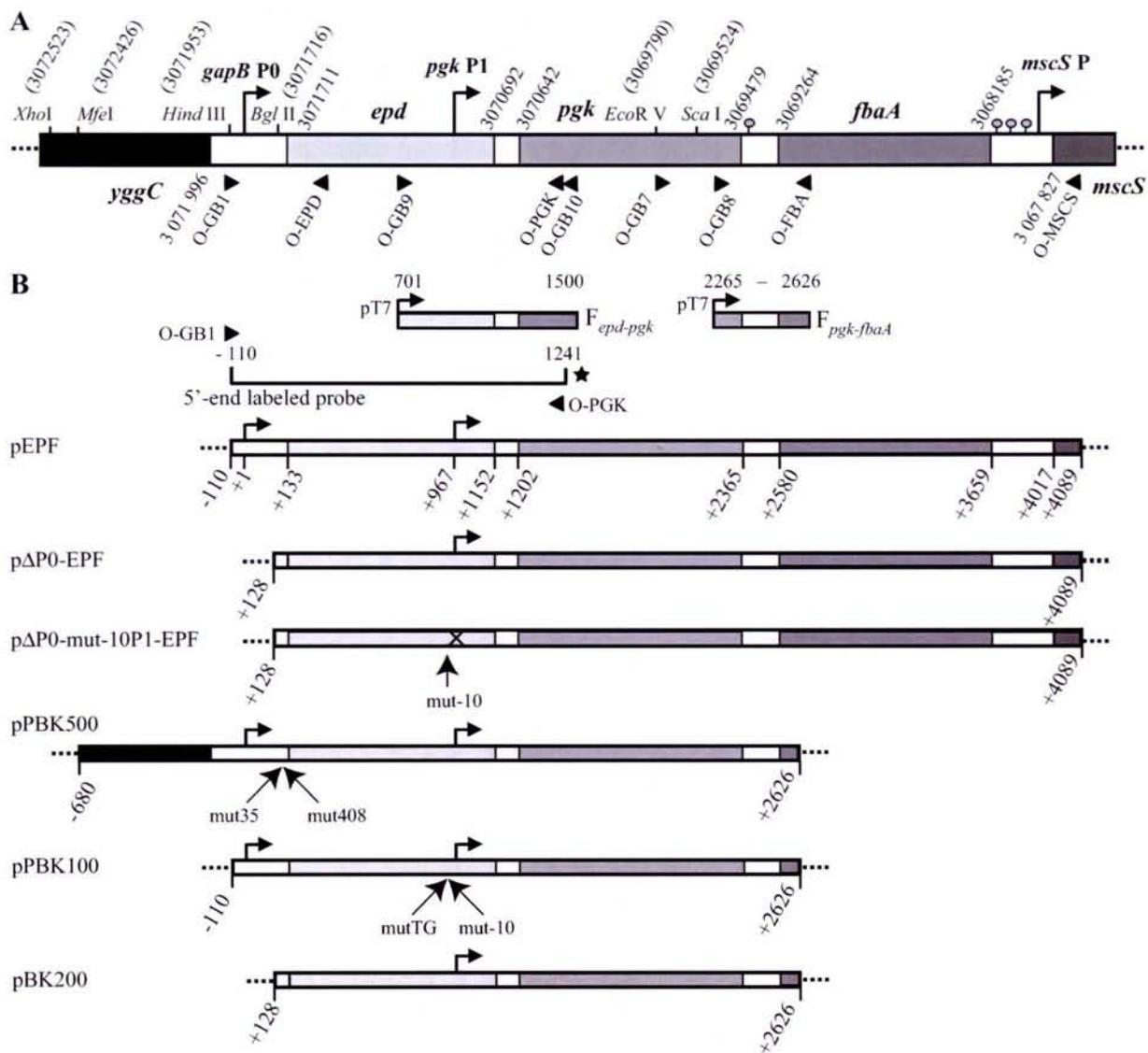


Figure 1
 Bardey *et al.*

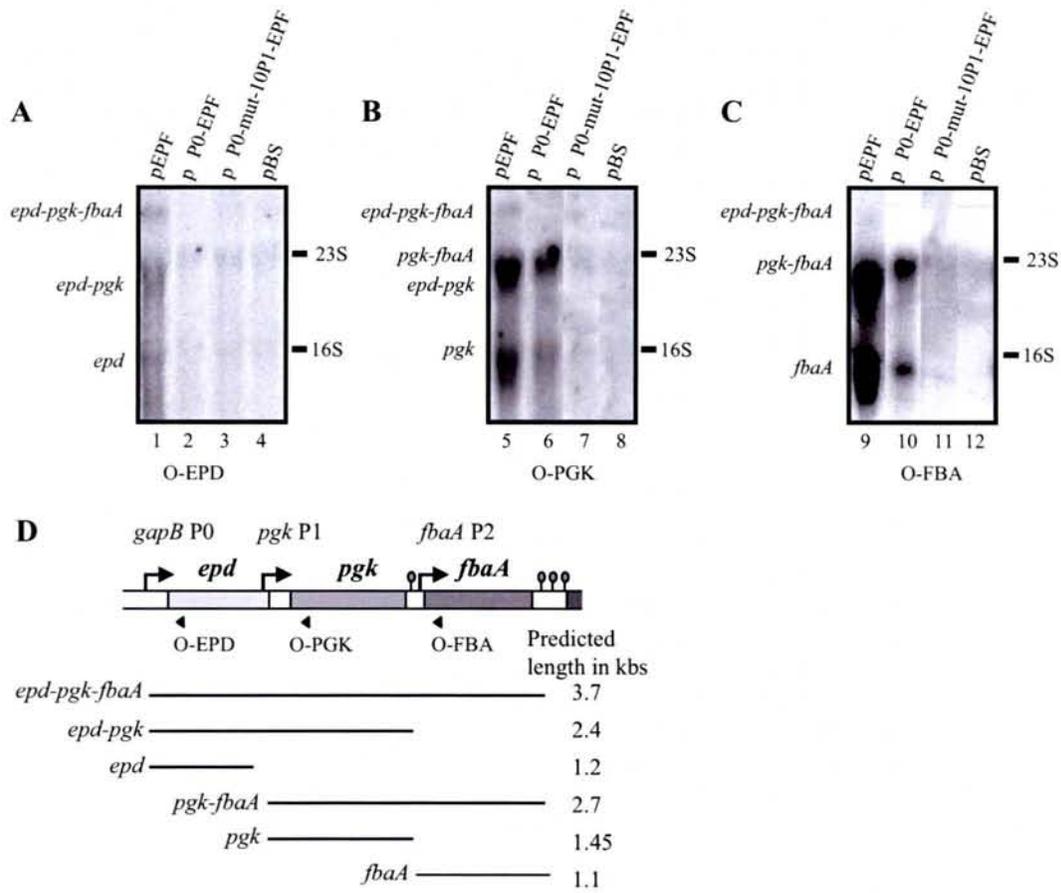


Figure 2
Bardey *et al.*

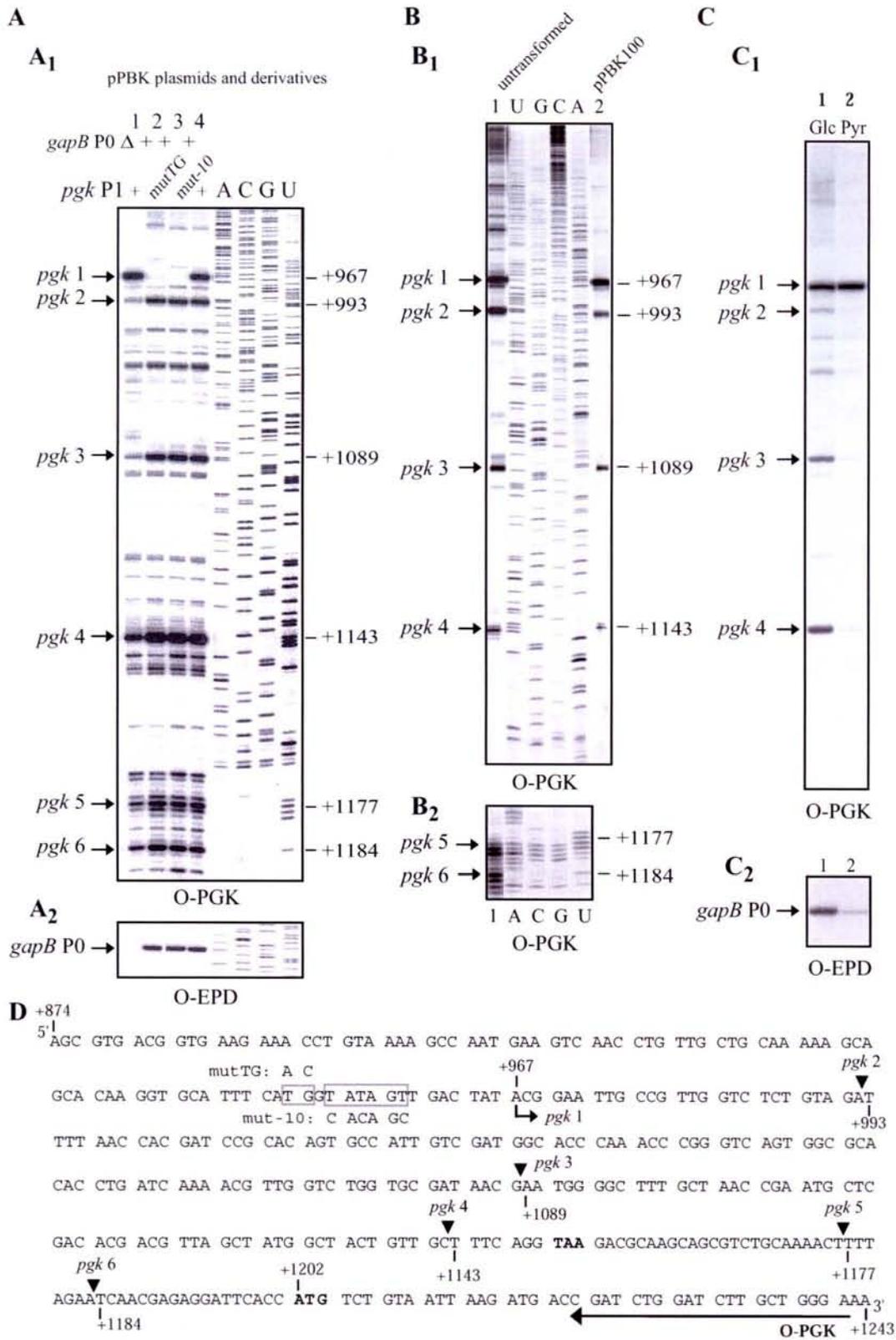


figure 3
 Bardey *et al.*

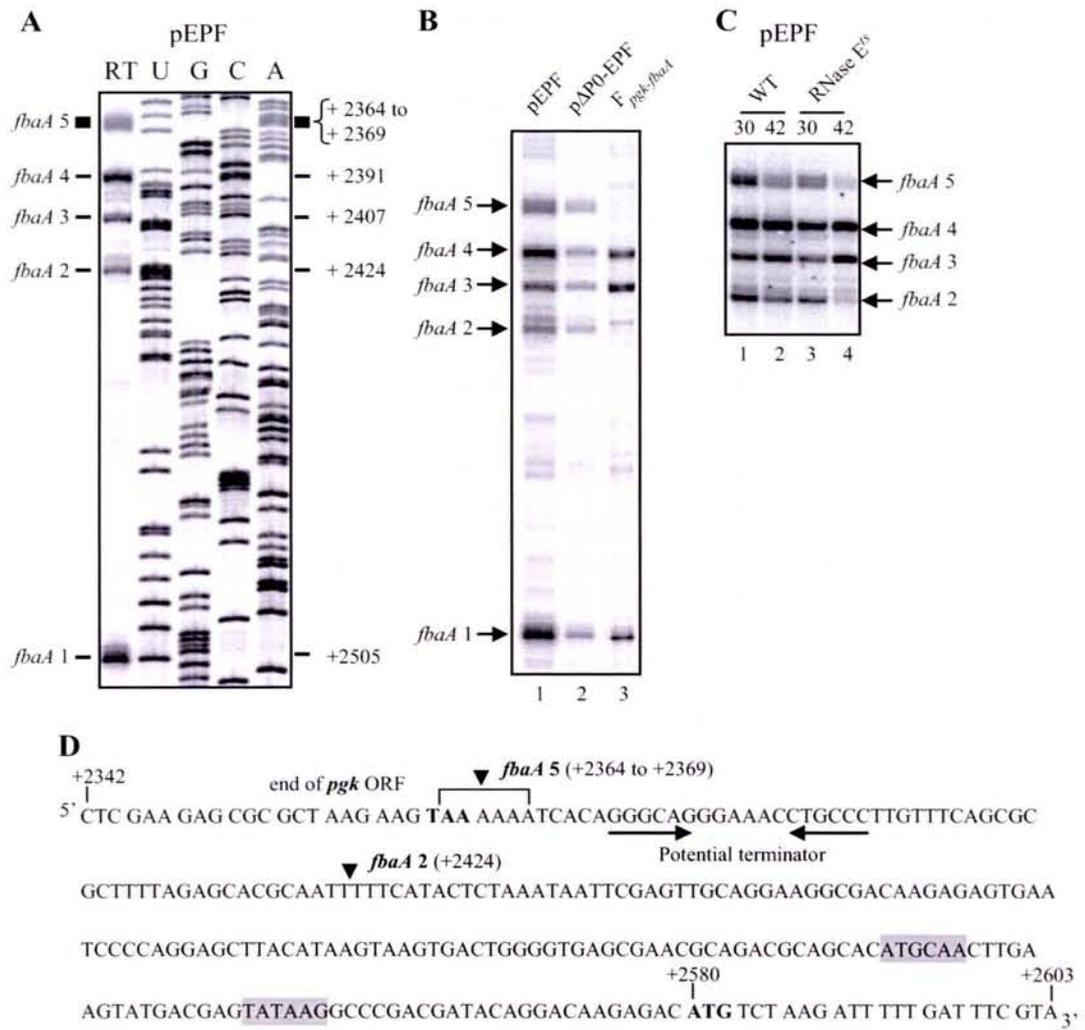


figure 4
 Bardey *et al.*

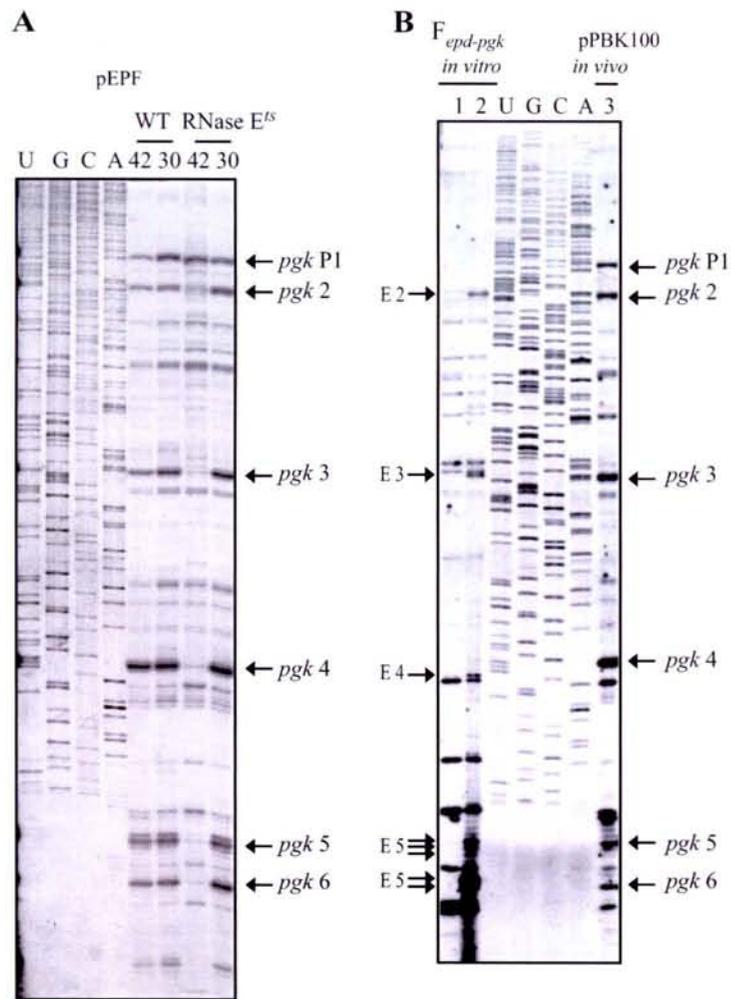


figure 5
Bardey *et al.*

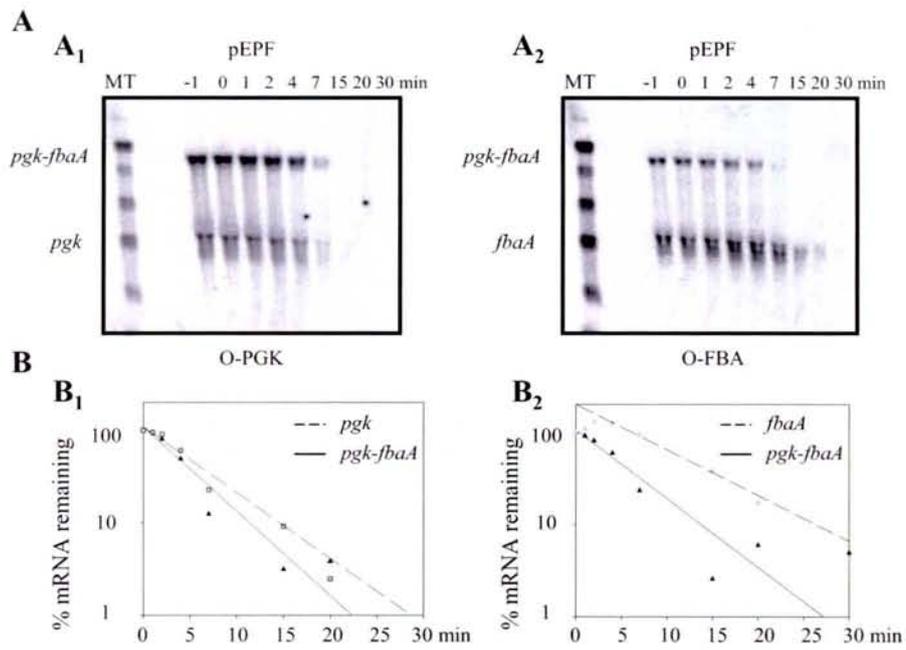


Figure 6
 Bardey *et al.*

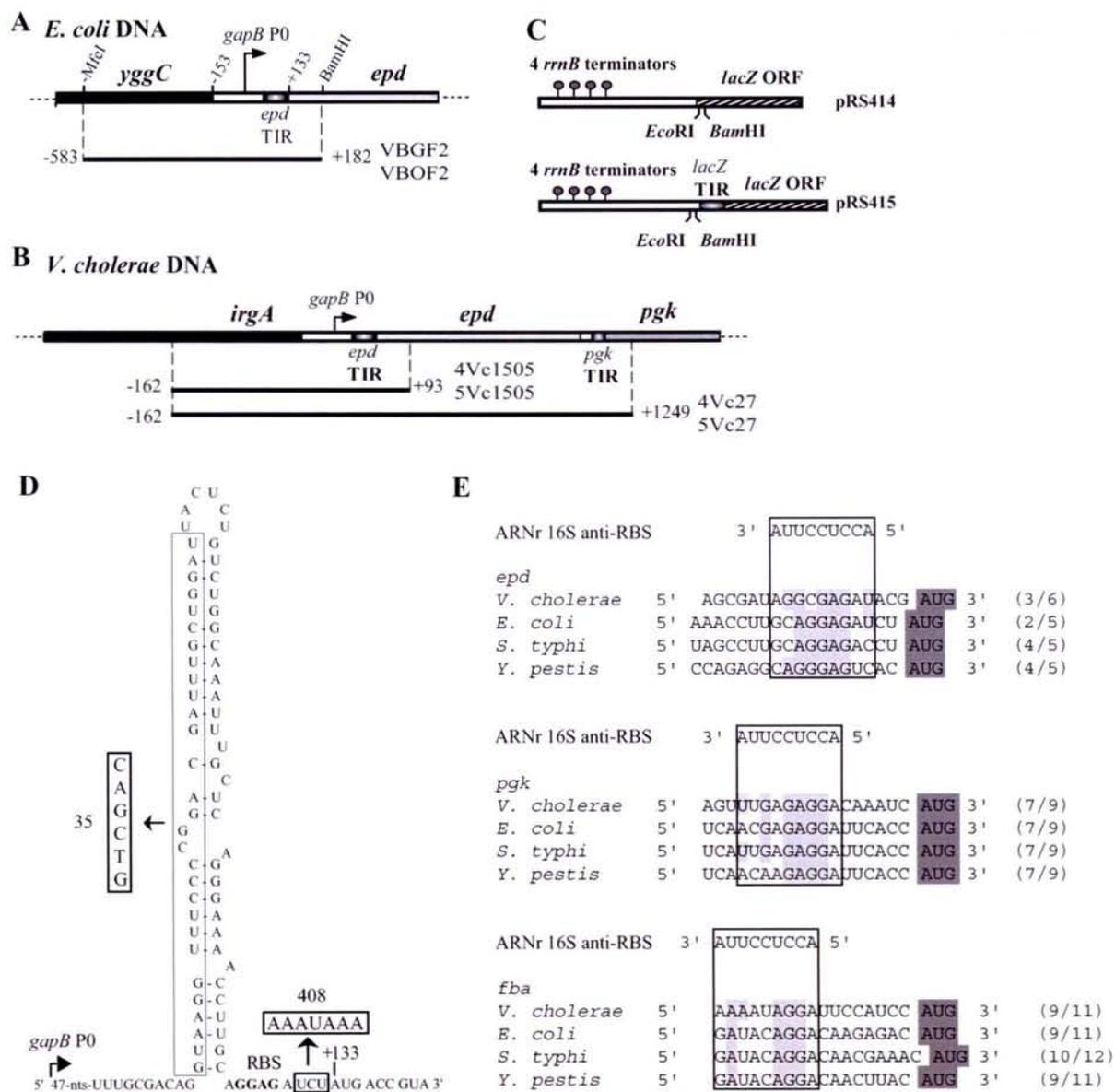


Figure 7
Bardey *et al.*

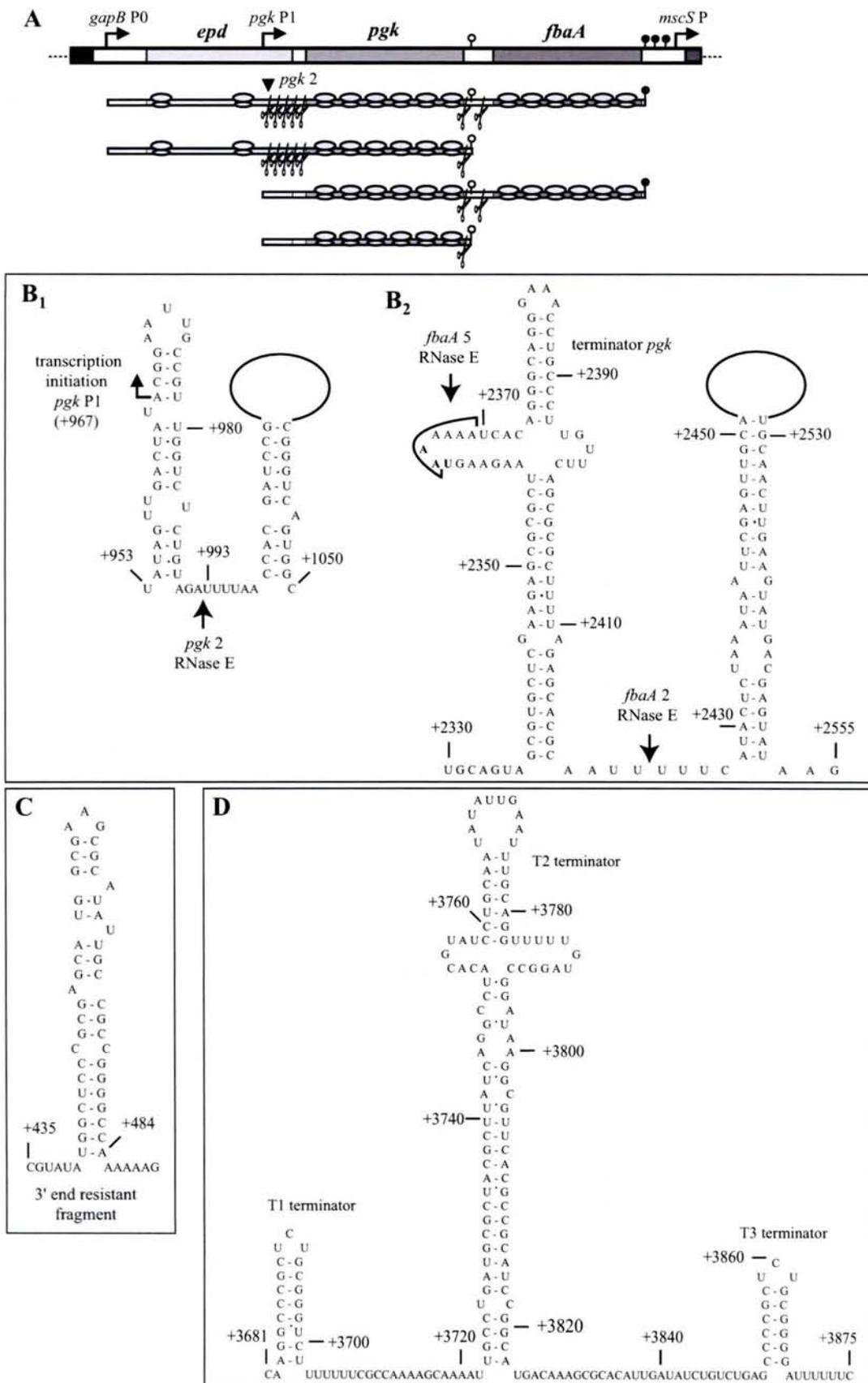


Figure 8
Bardey et al.

III.) Travaux complémentaires réalisés en vue de déterminer si des évènements post-transcriptionnels interviennent aussi dans la régulation de l'expression du gène *gapA*.

III.1) Test de l'effet du glucose sur les promoteurs *gapA* P1 et *gapB* P0 par emploi de fusions transcriptionnelles *gapA* P1-*lacZ* et *gapB* P0-*lacZ*.

Les travaux décrits dans l'article 2 montraient l'importance des étapes post-transcriptionnelles dans la régulation de l'expression du triplet d'ORF *epd-pgk-fbaA*. On pouvait donc se demander s'il en était de même pour le gène *gapA*. Pour cela, nous avons réalisé une fusion transcriptionnelle entre le gène *gapA* et l'ORF *lacZ*. L'ORF *lacZ* et sa région d'initiation de la traduction ont été placées en aval du site d'initiation de la transcription du promoteur *gapA* P1 (position +27 par rapport au site d'initiation *gapA* P1) (Fig. 69 A). Pour cela, le fragment d'ADN allant des positions -380 à +27 par rapport au site d'initiation *gapA* P1 a été amplifié par PCR avec les oligonucléotides 1429 et 1670, introduisant, respectivement, des sites de restriction *Hind*III et *Eco*RI. Après hydrolyse par *Hind*III, l'extrémité cohésive générée a été transformée en extrémité à bout franc, par une réaction de remplissage avec la T7 ADN pol. Ce fragment a ensuite été hydrolysé par l'enzyme *Eco*RI, puis cloné dans le vecteur pRS415 hydrolysé par les enzymes de restriction *Eco*RI et *Sma*I.

Cette construction a été placée à l'état de copie unique dans le chromosome de la souche TG1 d'*E. coli*, ceci en utilisant le phage λ . La méthode étant décrite dans la partie Matériels et Méthodes, nous n'en donnerons que le principe. Une souche d'*E. coli* a été transformée par le plasmide portant la fusion transcriptionnelle, puis a été infectée par le phage λ RS45. Nous avons ensuite sélectionné des phages infectieux recombinants qui portaient la fusion transcriptionnelle. Nous avons enfin infecté les cellules de la souche *E. coli* TG1 par ces phages, puis sélectionné les cellules où le phage portant la fusion transcriptionnelle s'était intégré à l'état de copie unique sur le chromosome au niveau du site *att*.

V. Bardey avait créé une fusion transcriptionnelle similaire en utilisant la région d'ADN renfermant le promoteur *gapB* P0. L'ADN du plasmide VBOF17c contenant la TIR de *lacZ* et de l'ORF *lacZ* était localisé à la position +127 par rapport au site d'initiation de la transcription du promoteur *gapB* P0 (Fig. 69 A).

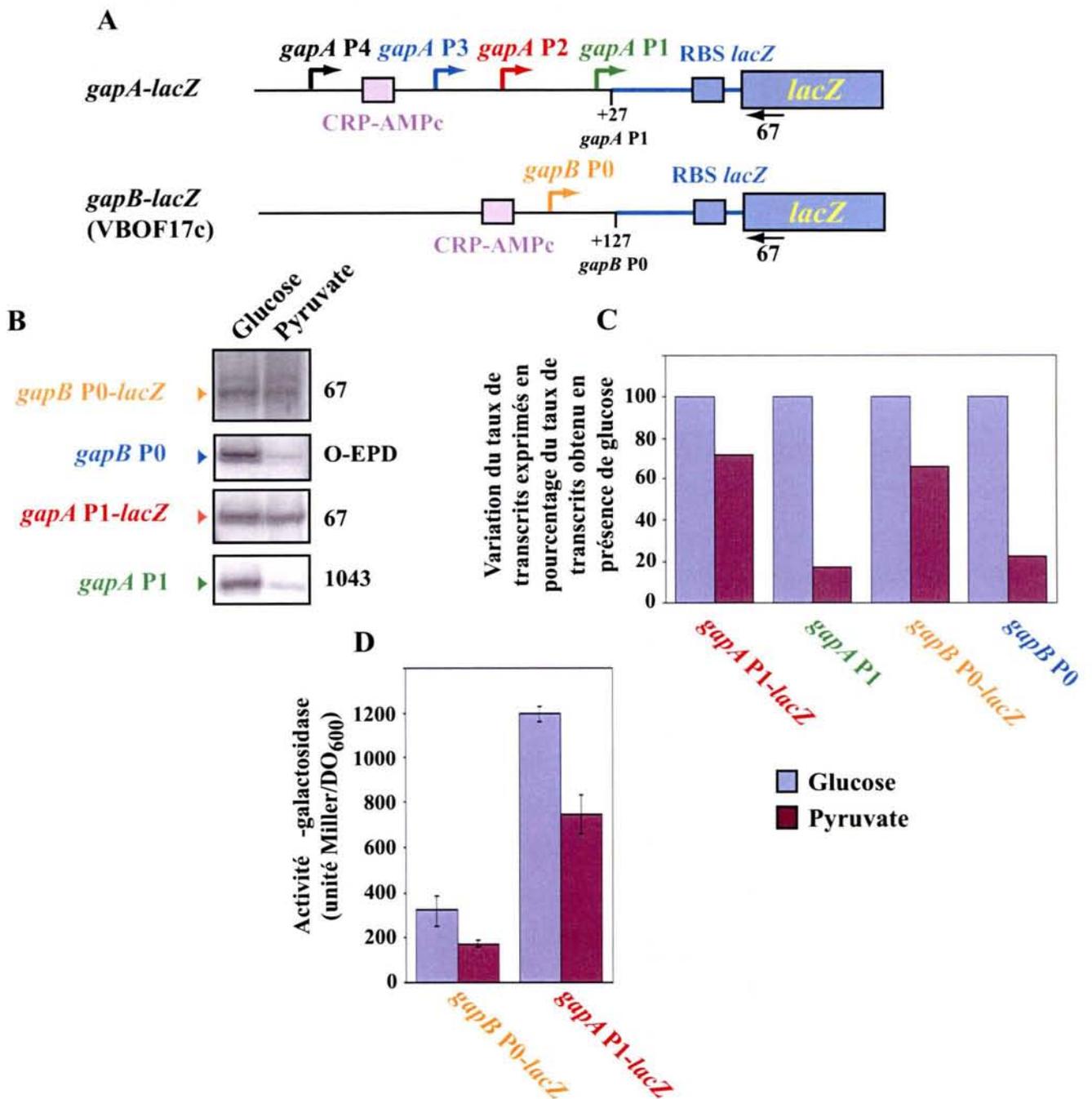


Figure 69 : L'augmentation du taux des transcrits *gapA P1* et *gapB P0* en présence de glucose dans le milieu de culture est lié à la fois à un effet d'ordre transcriptionnel et à un effet d'ordre post-transcriptionnel.

A) Représentation schématique des fusions transcriptionnelles *gapA-lacZ* et *gapB-lacZ* (VBOF17c ; V. Bardey) utilisées dans cette expérience. Les régions d'ADN contenant les différents promoteurs sont indiquées en noir, alors que les régions d'ADN du plasmide pRS415 sont indiquées en bleu. La séquence du site de fixation du ribosome (RBS) et l'ORF *lacZ* sont indiquées en bleu. Le site de fixation du complexe CRP-AMPc est indiqué en violet. L'oligonucléotide (67) qui a servi d'amorce pour les analyses par élongation d'amorce est indiqué.

B) Analyse par extension d'amorce de 20 µg d'ARN totaux extraits de cellules monolysogènes TG1 contenant les constructions *gapA-lacZ* et *gapB-lacZ*. Ces cellules avaient été cultivées en présence de glucose ou de pyruvate jusqu'à une DO_{600nm} de 0,5. Les extensions d'amorces ont été réalisées avec les oligonucléotides 1043, O-EPD et 67, complémentaires, respectivement, des séquences des ARNm *gapA*, *epd* et *lacZ*. L'oligonucléotide utilisé pour l'extension est indiqué à droite des autoradiogrammes.

C) Quantification des variations du taux de transcrits en fonction de la source de carbone. Pour chaque transcrit, la valeur obtenue en présence de pyruvate a été exprimée en pourcentage de celle obtenue en présence de glucose prise comme référence (100%).

D) Les souches TG1 monolysogènes ont été cultivées en présence de glucose ou en présence de pyruvate. L'activité -galactosidase a été mesurée en phase exponentielle de croissance. Les valeurs indiquées sont la moyenne des valeurs obtenues dans deux expériences distinctes.

L'utilisation des souches monolysogènes contenant chacune de ces constructions devait nous permettre de différencier les effets activateurs de la présence de glucose dans le milieu de culture vis-à-vis des promoteurs *gapA* P1 et *gapB* P0, des effets combinés de ce même glucose sur l'activité de ces promoteurs et sur les étapes post-transcriptionnelles qui co-existent lorsque le taux d'expression du gène *gapA* ou du cluster de gènes *epd-pgk-fbaA* est étudié *in vivo*.

Les deux souches d'*E. coli* TG1 comportant les constructions *gapA-lacZ* et *gapB-lacZ* à l'état de monocopie ont été cultivées en milieu minimum M63, en présence de 0,2 % de glucose ou 0,1 % de pyruvate comme seule source de carbone. Les cellules ont été collectées lorsque la DO_{600nm} avait atteint la valeur de 0,5. Les cellules ont alors été fractionnées en deux lots, l'un a servi à mesurer l'activité β -galactosidase présente dans le sonicat (cf. article n°2), l'autre a été utilisé pour la préparation des ARN totaux, au sein desquels nous avons estimé les taux relatifs des transcrits endogènes et des transcrits produits par les fusions transcriptionnelles. Ainsi, dans la souche transformée par la fusion *gapB* P0-*lacZ*, nous avons comparé les variations d'expression du transcrit fusionné *gapB* P0-*lacZ* et du transcrit endogène produit à partir du promoteur *gapB* P0. De même, pour la souche comportant la construction *gapA* P1-*lacZ*, nous avons comparé les variations d'expression du transcrit fusionné *gapA* P1-*lacZ* et du transcrit endogène produit à partir du gène sauvage *gapA*.

Ces variations de taux d'expression ont été estimées par la technique d'élongation d'amorce en utilisant, respectivement, les oligonucléotides 1043 (article n°1), O-EPD (article n°2) et 67 (Fig. 69 A) pour l'étude des ARNm *gapA* sauvage, *epd* sauvage et des ARN fusionnés *gapA* P1-*lacZ* et *gapB* P0-*lacZ* (le protocole utilisé était le même que celui décrit dans l'article n°1). Il faut noter que l'oligonucléotide O-EPD utilisé pour l'analyse des produits de transcription à partir du promoteur *gapB* P0 endogène est complémentaire au segment 5' terminal résistant à l'action des RNases (article n°2).

Les résultats obtenus pour la fusion *gapB-lacZ* montrent qu'en présence de glucose, le taux d'ARNm initiés au promoteur *gapB* P0-*lacZ* est augmenté d'un facteur 1,5, alors que le taux d'ARNm initiés au promoteur *gapB* P0 endogène est augmenté d'un facteur 4,4 (Fig. 69 B et C). Il est à noter que l'activité β -galactosidase exprimée à partir de la fusion *gapB* P0-*lacZ* augmente d'un facteur 1,8 lorsque la culture est réalisée en présence de glucose (Fig. 69 D). Ces données qui confirment des résultats préliminaires qu'avait obtenus V. Bardey indiquent que la présence de l'ORF entière *epd* en aval du promoteur *gapB* P0 a une influence sur le taux d'ARN initiés au promoteur *gapB* P0 présent dans les cellules. Une partie de l'effet

activateur du glucose sur le taux de transcrits *gapB* P0 semble donc liée à un effet post-transcriptionnel qui pourrait concerner la stabilité et/ou l'efficacité de traduction de l'ORF *epd*. L'autre partie de l'effet glucose (soit environ 30 % de l'effet total) serait lié directement aux variations d'efficacité d'initiation de la transcription du promoteur *gapB* P0, puisque cet effet est retrouvé pour la fusion transcriptionnelle.

Nous avons obtenu des résultats similaires pour la fusion *gapA-lacZ*. En effet, le taux d'ARNm initiés au promoteur *gapA-lacZ* P1 de la fusion *gapA-lacZ* est augmenté d'un facteur 1,4 en présence de glucose, alors que le taux de transcrits initiés au promoteur *gapA* P1 endogène est augmenté d'un facteur 5,7 (Fig. 69 B et C). De la même manière, le taux d'activité β -galactosidase produit à partir de la fusion *gapA-lacZ* est augmenté d'un facteur 1,6 (Fig. 69 D), ce qui montre que l'effet du glucose sur le taux de transcrits *gapA* P1 est aussi à la fois lié à un mécanisme d'ordre transcriptionnel et à un ou plusieurs mécanismes post-transcriptionnels.

Ces résultats suggèrent que les régulations observées au niveau de ces deux gènes en présence de glucose font intervenir un effet lié aux caractéristiques des promoteurs et un effet lié à la stabilité et/ou la régulation de la traduction des transcrits.

III.2) Discussion et perspectives des résultats obtenus.

En ce qui concerne le promoteur *gapB* P0, on pourrait imaginer que deux paramètres puissent intervenir dans l'effet observé de la présence de glucose sur les étapes post-transcriptionnelles :

- 1) un changement de l'efficacité de traduction de l'ORF *epd* qui, comme nous l'avons vu, est très faible et limitée par des propriétés de la TIR et, vraisemblablement un régulateur agissant *en trans*.
- 2) un changement de la stabilité du transcrit qui, comme nous l'avons vu dans l'article 2, subit un processus complexe de clivage.

Un changement de l'efficacité de traduction paraît peu probable car, quel que soit le milieu de culture utilisé le taux de production d'E4PDH reste très faible. Il est donc plus vraisemblable que l'effet positif du glucose observé se situe au niveau de la stabilité du transcrit. Il est en effet possible que l'activité RNase E soit plus limitée en présence de glucose dans le milieu de culture. En lien avec cette hypothèse, il a été montré que l'activité RNase E peut être régulée par les conditions de croissance, des effets de la température et de la présence d'oxygène ayant déjà été montrés (Georgellis *et al.*, 1993). De plus, un facteur appelé RraA limite

l'activité de la RNase E en se fixant au niveau de son extrémité C-terminale (Lee *et al.*, 2003). Le taux de ce facteur pourrait augmenter en présence de glucose, ce qui n'a pas encore été étudié. Une autre explication possible pourrait être, qu'en présence de glucose dans le milieu de culture, la vitesse de croissance des cellules est plus élevée, ce qui conduit à une augmentation de la transcription des gènes des opérons ribosomiques. Dans ces conditions, la RNase E est vraisemblablement mobilisée au niveau de ces ARN (notamment pour la maturation de l'ARNr 5S), ce qui pourrait limiter son action au niveau des ARNm comme les transcrits *epd-pgk-fbaA*. Ces transcrits seraient ainsi moins efficacement hydrolysés en présence de glucose.

Un moyen de tester ces hypothèses sera de réaliser une construction dans laquelle la région *epd-pgk-fbaA* est placée sous contrôle d'un autre promoteur, non inductible par le glucose. Les produits d'expression de cette construction pourraient alors être étudiés pour différentes conditions de culture et dans des souches exprimant et n'exprimant pas le facteur RraA, ceci en présence de glucose ou de pyruvate dans le milieu de culture.

En ce qui concerne le gène *gapA*, aucune étude n'avait été faite sur l'existence possible de clivages de l'ARNm *in vivo*. Nous ne pouvions donc pas savoir si l'effet observé du glucose sur les étapes post-transcriptionnelles était lié à des changements de stabilité ou de l'efficacité de traduction de l'ARNm *gapA* P1.

Nous avons donc dans un premier temps utilisé les approches par empreinte à la nucléase S1 et extension d'amorce pour essayer de savoir si des produits de clivage de l'ARNm *gapA* P1 étaient détectables *in vivo*. Nous n'avons détecté aucun produit de clivage stable, donc le processus de dégradation du messenger doit conduire à sa disparition totale. Il faudrait maintenant tester, par l'approche rifampicine utilisée dans l'article n°2, si cet ARNm est plus stable en présence de glucose. Toujours dans l'idée que l'ARNm *gapA* est stabilisé en présence de glucose, on peut envisager que cette augmentation de stabilité soit liée à une augmentation de l'efficacité de traduction en présence de glucose. Nous nous sommes alors demandé si l'expression du gène *gapA* n'était pas soumise à un contrôle par le système CsrA/CsrB ou C.

III.3) Le système CsrA/BC n'est pas responsable de l'effet post-transcriptionnel observé sur le promoteur *gapA* P1.

III.3.1) Présentation du système CsrA/BC.

Le système Csr (Carbon storage regulation) permet en effet la modulation de l'expression de gènes qui codent des protéines impliquées dans le métabolisme du carbone, via un contrôle post-transcriptionnel. La protéine CsrA est connue pour exercer un contrôle négatif sur la synthèse de glycogène en phase stationnaire de croissance et, au contraire, pour avoir un rôle positif sur la glycolyse en stimulant l'expression des gènes *pgi* codant la phosphoglucose isomérase, *tpi*, la triose phosphate isomérase, *pfkA*, la phosphofructokinase et *eno*, l'énolase (pour revue : Romeo, 1998).

Le mode d'action de la protéine CsrA a été étudié au niveau du gène *glgC*, qui code l'ADP glucose pyrophosphorylase, une enzyme impliquée dans le métabolisme du glycogène. La protéine CsrA se fixe au niveau de la séquence de fixation du ribosome de l'ARNm *glgC* et empêche la fixation de la sous-unité 30S, et donc l'initiation de la traduction de l'ARNm (Baker *et al.*, 2002). Dans le cas d'une régulation positive par CsrA, comme par exemple au niveau de l'ARNm *flhDC*, qui code le facteur transcriptionnel FlhD₂C₂ contrôlant la synthèse des flagelles, la protéine se fixe au niveau de la séquence amont non traduite et stabilise ce transcrite (Wei *et al.*, 2001).

L'activité de CsrA est régulée par deux ARN régulateurs appelés CsrB et CsrC qui possèdent un grand nombre de sites de fixation pour CsrA (18 pour CsrB et 9 pour CsrC) (Weilbacher *et al.*, 2003). La fixation de CsrA sur ces ARN piège cette protéine qui ne peut alors plus se fixer sur ces ARNm cibles et jouer son rôle de régulateur post-transcriptionnel (Gudapaty *et al.*, 2001). Il est à noter que la quantité d'ARN CsrB présent dans la cellule ne semble pouvoir séquestrer qu'environ 30 % de la protéine CsrA présente dans la cellule, le second, l'ARN CsrC, qui a un rôle équivalent pourrait aussi séquestrer environ 30 % de cette protéine (Gudapaty *et al.*, 2001; Weilbacher *et al.*, 2003).

III.3.2) Etude expérimentale de l'effet potentiel de CsrA et CsrB sur l'expression du gène *gapA*.

Pour tester si le taux de transcrits initiés au promoteur *gapA* P1 est régulé par le système Csr, nous avons, dans un premier temps, réalisé deux constructions dérivées du plasmide

multicopie (pUC19), qui permettent l'une la surexpression de la protéine CsrA (pUC::*csrA*), l'autre la surproduction de l'ARN CsrB (pUC::*csrB*). Les cellules XL1 blue ont été transformées par ces deux constructions et nous avons cultivé ces cellules transformées en milieu LB additionné d'ampicilline jusqu'à une DO_{600nm} de 0,5, puis nous avons extrait les ARN totaux. Nous avons vérifié que la protéine CsrA et l'ARN CsrB étaient surexprimés en testant leurs effets sur les taux d'ARNm *glgC* et *pfkA* par élongation d'amorce par la transcriptase inverse, ceci en utilisant les amorces 2318 et 2319 complémentaires des ORF de ces ARNm (Fig. 70). Comme attendu, en présence de la protéine CsrA, le taux d'ARNm *pfkA* est augmenté et lorsque l'ARN CsrB est surexprimé, le taux d'ARNm *glgC* est légèrement augmenté. Par contre, l'effet de la surproduction de la protéine CsrA sur le taux de transcrits initiés au promoteur *gapA* P1 est faible (Fig. 70). On note une légère augmentation (facteur de 1,2). Nous observons très peu d'effet de la surproduction de l'ARN CsrB sur le taux de transcrits *gapA* P1. Par contre, une augmentation du taux de transcrits initiés au promoteur *gapA* P2 est observé (Fig. 70), ce qui pourrait refléter la création d'un état de stress de la cellule lorsque l'ARN CsrB est surexprimé.

Le système Csr ne semblait donc pas être impliqué de manière importante dans la régulation post-transcriptionnelle du taux de transcrits initiés au promoteur *gapA* P1. D'une manière générale, la régulation de l'expression des gènes par le système CsrA/CsrB/CsrC est surtout importante en phase stationnaire de croissance. Or les effets que nous avons observés ont lieu en phase exponentielle de croissance. Une possibilité serait qu'un autre ARNnc soit impliqué dans la régulation du taux de traduction et donc de la stabilité de l'ARNm *gapA* P1.

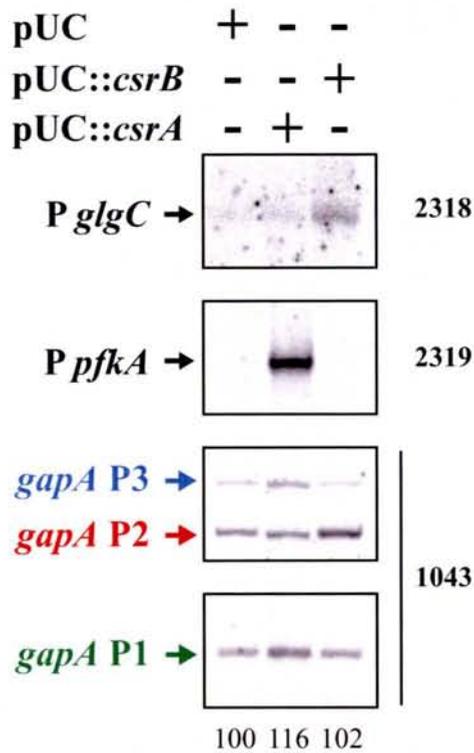


Figure 70 : Recherche d'un rôle éventuel du système CsrA/CsrB sur l'expression du gène *gapA*. Analyse par extension d'amorce par la transcriptase inverse des ARN extraits des souches d'*E. coli* XL1 blue transformées par le plasmide pUC::*csrA*, surproduisant la protéine CsrA d'*E. coli*, par le plasmide pUC::*csrB*, surproduisant l'ARN non codant CsrB ou par le plasmide pUC sans insert. Les cellules ont été cultivées en milieu LB additionné d'ampicilline, jusqu'à une DO_{600nm} de 0,5. Les ADNc ont été synthétisés en utilisant des oligonucléotides complémentaires aux transcrits *gapA* (1043), *pfkA* qui codent la phosphofruktokinase (2319) et *glgC* qui codent l'ADP glucose pyrophosphorylase (2318). Ils ont été fractionnés par électrophorèse en gel dénaturant. L'identité du plasmide utilisé pour transformer les souches est indiquée au-dessus des pistes. La quantification de la radioactivité présente dans la bande correspondant au transcrit *gapA* P1 est indiquée en-dessous des autoradiogrammes. Le taux de transcrits *gapA* P1 obtenu pour le plasmide pUC est pris comme référence (100%).

Partie 4 : Analyse de l'effet de l'absence du transporteur spécifique du glucose EIIBC^{Glc} sur le transcriptome d'*E. coli*

I.) Présentation de l'article n°3.

I.1) **Présentation des objectifs.**

Nous avons constaté qu'une part de l'augmentation du taux de transcrits *gapA* P1 et *gapB* P0 en présence de glucose dans le milieu de culture était d'ordre transcriptionnel (cf. partie précédente). Par ailleurs, l'absence de cette augmentation du taux de transcrits *gapA* P1 et *gapB* P0 en absence d'expression de protéine EIIBC^{Glc} active (Charpentier *et al.*, 1998) montrait un lien entre la présence de cette protéine et l'activation des promoteurs *gapA* P1 et *gapB* P0. La question était de savoir par quel mécanisme EIIBC^{Glc} peut entraîner une activation de ces deux promoteurs. Chez *E. coli*, d'autres promoteurs sont connus pour avoir une augmentation de leur activité en présence de glucose dans le milieu de culture dépendante de l'expression de la protéine EIIBC^{Glc}. C'est le cas pour le promoteur *ptsH* P0 permettant la transcription de l'opéron *pts* et le promoteur *ptsG* P1 du gène codant la protéine EIIBC^{Glc} (cf. Introduction). Dans ce cas, l'activation repose sur l'action de la protéine Mlc (cf. Introduction (Fig. 30) et pour revue : Plumbridge, 2002). En effet, en présence de glucose, l'inhibiteur transcriptionnel Mlc est séquestré au niveau de la membrane par EIIBC^{Glc} et ne peut plus réprimer l'initiation de la transcription au niveau de ces deux promoteurs (Lee *et al.*, 2000; Tanaka *et al.*, 2000; Nam *et al.*, 2001). La protéine Mlc a aussi un rôle de répresseur de la transcription des gènes *malt* (Decker *et al.*, 1998) et *manXYZ* (Plumbridge, 1998a), dont les produits seraient liés au métabolisme du glucose. Une possibilité était donc que les promoteurs *gapA* P1 et *gapB* P0 soient activés en présence de glucose par le même mécanisme. Nous avons donc testé l'implication possible de Mlc sur l'activité de ces deux promoteurs. L'examen de la séquence des deux régions promoteurs ne mettait pas en évidence de sites potentiels de fixation de Mlc, ce qui suggérait que Mlc n'interagit pas directement avec ces promoteurs. Une action indirecte pouvait être possible. Nous avons confirmé expérimentalement que ce n'était pas le cas en utilisant une souche *mlc*⁻, ne produisant plus la protéine Mlc (don de J. Plumbridge, Paris). En effet, nous n'avons pas détecté de changement significatif du taux d'initiation de la transcription aux promoteurs *gapA* P1 (Fig. 1 article n°3)

et *gapB* P0 (données non montrées) dans cette souche, ni de plus faible augmentation du taux de transcrits correspondants en présence de glucose. Nous avons donc conclu que l'activation des promoteurs *gapA* P1 et *gapB* P0 en présence de glucose relève d'un autre mécanisme, indépendant de la protéine Mlc.

I.2) Présentation de la stratégie.

Nous avons alors pensé qu'un moyen d'obtenir des informations sur ce mécanisme serait d'étudier l'effet global de l'absence d'expression de la protéine EIIBC^{Glc} active sur la transcription des gènes chez *E. coli*. Nous pouvions ainsi espérer avoir des informations soit sur des gènes dont l'expression serait diminuée et dont les produits pourraient être impliqués dans l'activation des promoteurs *gapA* P1 et *gapB* P0, soit sur des gènes soumis à une régulation similaire à celle à laquelle les gènes *gapA* P1 et *gapB* P0 sont soumis. Pour cela, nous avons utilisé une approche puces à ADN. Des membranes de nylon, portant l'ensemble des ORF d'*E. coli* amplifiées par PCR, nous ont été fournies dans le cadre d'un consortium de laboratoires français ayant obtenu un soutien de la part du ministère et du CNRS pour développer le programme « Le transcriptome d'*Escherichia coli* : analyse fonctionnelle globale d'une bactérie modèle ». Le responsable scientifique du projet était P. Bouloc (Orsay). Les membranes de nylon portant les ORF ont été fabriquées par la société Sigma Genosys. L'ensemble du génome était couvert par 3 membranes comportant pour les deux premières, l'ADN de 1520 ORF distinctes et pour la troisième, l'ADN de 1250 ORF. Chaque ORF était déposée en double. Par ailleurs, de l'ADN génomique total d'*E. coli* était déposé comme témoin sur chaque membrane.

Nous avons utilisé ces membranes pour comparer les taux relatifs des ARNm extraits de la souche d'*E. coli* TP2504 *ptsG*⁻ (*ptsG22*) produisant une protéine EIIBC^{Glc} inactive et de la souche isogénique TP2503 *ptsG*⁺ (De Reuse et Danchin, 1988) (Fig. 71). Les hybridations ont ensuite été réalisées avec des ADNc radioactifs (³³P) synthétisés à partir des ARN totaux extraits de cellules des souches TP2503 et TP2504. Les cellules ont été cultivées en milieu minimum M63 en présence de glucose à 0,2 % et récoltées lorsque la DO_{600nm} avait atteint la valeur de 0,5. Les membranes ont ensuite été analysées avec un PhosporImager et les images ont été traitées avec le logiciel ArrayVision (Imaging Research Inc.). La normalisation des valeurs a été effectuée avec la fonction « overall-spot normalisation » du logiciel. Les résultats ont ensuite été intégrés dans un tableau fourni par Sigma Genosys à l'aide du logiciel Excel. Seuls les gènes dont l'expression était augmentée ou diminuée d'un facteur supérieur

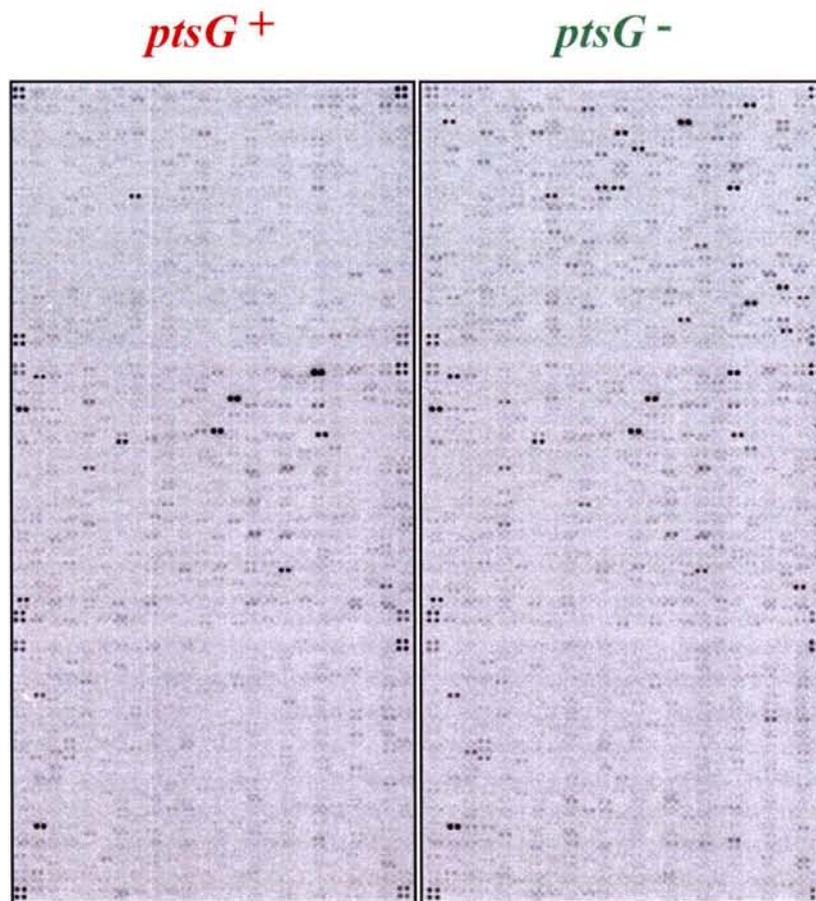


Figure 71 : Résultats obtenus après hybridation des puces à ADN Sigma-Genosys (membrane de nylon comportant des fragments d'ADN correspondant aux 4290 ORFs d'*E. coli*). Les ARN totaux ont été extraits des cellules d'*E. coli ptsG*⁺ (TP2503) et *ptsG*⁻ (TP2504) cultivées en milieu M63 glucose et récoltées en phase exponentielle de croissance. Les ARN totaux ont ensuite été extraits et les ADNc produits par transcription inverse réalisée sur 40 µg d'ARN totaux ont été marqués par du dCTP [α -³³P].

ou égal à 2 dans la souche *ptsG⁻* par rapport à la souche *ptsG⁺*, ceci dans les trois expériences indépendantes que nous avons réalisées, ont été pris en considération.

I.3) Résumé des résultats marquants obtenus.

Le tableau indiquant les gènes dont l'expression varie en fonction de la souche considérée est présenté dans la publication (tableau 1). Il apparaît que l'expression des gènes de trois types de composants cellulaires varie fortement en fonction du contexte génétique, il s'agit de gènes codant :

- des constituants de la membrane externe : porines et flagelles,
- des enzymes impliquées dans la détoxification du milieu,
- des enzymes impliquées dans le catabolisme des composés carbonés.

Malheureusement, l'approche ne permettait de détecter que de très fortes variations du taux de transcrits. Nous n'avons pas détecté de variation forte du taux de transcrits initiés pour la région promoteurs *gapA* entre les deux souches, ce qui peut s'expliquer par la multiplicité des promoteurs. De même, nous n'avons pas détecté de variation pour le taux de transcrits initiés aux promoteurs *gapB* P0 et *ptsH* P0. De ce fait, nous avons focalisé nos efforts sur l'étude des gènes dont le taux d'expression était très différent dans les deux souches dans les conditions de croissance utilisées et plus particulièrement sur des gènes pour lesquels des informations étaient disponibles dans la littérature.

Dans un premier temps, nous avons essayé de confirmer les résultats obtenus par l'approche puces à ADN par une analyse réalisée par extension d'amorce, ceci pour les ARNm codant les porines OmpF et OmpC, et les ARNm codant le facteur σ^{28} (FliA) et la flagelline FliC. Les résultats obtenus confirmaient que le taux d'ARNm *ompF*, *fliC* et *fliA* augmente en absence de la protéine EIIBC^{Glc} active, tandis que celui des ARNm *ompC* diminue (Fig. 2 dans l'article n°3). Il était important de vérifier que les variations observées au niveau ARN étaient bien corrélées aux variations du taux de protéine. Les ARNm *ompC* et *ompF* codent les porines du même nom dont les propriétés sont décrites dans la partie Introduction. Pour extraire ces protéines membranaires, nous avons utilisé une technique déjà décrite dans la littérature (Liu et Ferenci, 1998), et leur taux a été estimé par électrophorèse en gel SDS-PAGE contenant de l'urée (cf. Matériel et Méthodes et Fig. 3 dans l'article n°3). Les résultats confirmaient bien que le taux de porine OmpC diminuait dans la souche *ptsG⁻*, alors que le taux de porine OmpF augmentait dans ce contexte. Il est à noter que le pore créé par OmpF

possède un diamètre plus important que celui créé par OmpC (cf. Introduction). Ce changement d'expression était donc supposé permettre l'entrée de plus grandes molécules dans les cellules.

Le gène *fliA* code le facteur σ^{28} , qui permet l'initiation de la transcription des gènes de classe III à laquelle appartient le gène *fliC* (cf. Introduction). La protéine FliC codée par ce gène est un des constituants de la partie externe du flagelle. Nous avons utilisé la mesure de la mobilité des souches dans un milieu solide contenant une faible concentration en gélose pour estimer le taux d'expression de la protéine FliC. Nous avons ainsi observé que les souches *ptsG⁻* ont une mobilité plus grande que les souches *ptsG⁺* (Fig. 3 dans l'article n°3).

Sur la base d'autres données de la littérature, nous avons essayé de construire un modèle de régulation (Fig. 5 article n°3) permettant d'expliquer les variations phénotypiques que nous avons observées entre les souches *ptsG⁻* et *ptsG⁺* et nous avons réalisé des expériences visant à apporter des arguments en faveur de ce modèle.

Le modèle repose sur le fait que, lors du transport du glucose à travers la membrane interne par son système PTS spécifique, le taux de phosphorylation de la protéine EIIA^{Glc} est diminué (cf. Introduction). Dans le cas de la souche *ptsG⁻*, le transporteur EIIBC^{Glc} n'est plus présent dans la membrane de la cellule. La concentration intracellulaire de la protéine EIIA^{Glc} sous sa forme phosphorylée est donc supposée augmenter, ce que nous avons vérifié en utilisant des anticorps anti EIIA^{Glc} (Krin *et al.*, 2002). L'adénylate cyclase étant régulée positivement par EIIA^{Glc}-P, ceci suggère que le taux de complexe CRP-AMPc augmente dans la souche *ptsG⁻*. En utilisant une souche *cya⁻*, ne produisant plus d'AMPc, et une souche *crr⁻*, ne produisant plus la protéine EIIA^{Glc}, nous avons démontré l'implication du complexe CRP-AMPc dans le mécanisme d'expression différentielle des porines OmpF et OmpC que nous avons mis en évidence.

II) Article n°3 : An analysis of the consequences of the absence of the EIIBC^{Glc} specific transporter of glucose on porin and flagellin expressions in *Escherichia coli*.

B. Thouvenot, C. Branlant and B. Charpentier.

Publication n°3

An analysis of the consequences of the absence of the EIIBC^{Glc}
specific transporter of glucose on porin and flagellin expressions in
Escherichia coli.

B. Thouvenot, C. Branlant et B. Charpentier

En préparation.

An analysis of the consequences of the absence of the EIIBC^{Glc} specific transporter of glucose on porin and flagellin expressions in *Escherichia coli*.

BENOIT THOUVENOT, CHRISTIANE BRANLANT and BRUNO CHARPENTIER*

Maturation des ARN et Enzymologie Moléculaire, UMR 7567 CNRS-UHP Nancy I, Boulevard des Aiguillettes, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy cedex, France.

* To whom correspondence should be addressed.

Dr. CHARPENTIER Bruno, MAEM, UMR 7567 CNRS-UHP Nancy I, Boulevard des Aiguillettes, Faculté des Sciences, BP 239, 54506 Vandoeuvre-les-Nancy cedex, France.

Tel (33) 3 83 68 43 16 - Fax (33) 3 83 68 43 07.

E-mail: bruno.charpentier@maem.uhp-nancy.fr

Keywords: *Escherichia coli*, *gapA*, Mlc, transcriptome analysis, *ptsG*, CRP-cAMP, porin, OmpR, flagella.

Abstract.

In *Escherichia coli*, protein EIIBC^{Glc}, encoded by the *ptsG* gene, allows the specific transport and phosphorylation of glucose. This protein of the glucose phosphoenolpyruvate-dependent phosphotransferase system (PTS) is involved in the regulation of several genes encoding components of sugar transport and metabolism, through the sequestration of the protein Mlc, a negative regulator of transcription. Here, by study of the regulation of promoter *gapA* P1, we show that the EIIBC^{Glc} protein may be required for activation of gene transcription without the involvement of the Mlc protein. In addition, by a comparison of the transcriptomes of a *ptsG*⁺ and a *ptsG*⁻ strain by use of macroarrays, we show that the absence of expression of an active EIIBC^{Glc} protein results in a strong modification of the expression of several genes, especially genes encoding the OmpF and OmpC porins, flagella constituents, enzymes involved in the metabolism of carbon sources, and detoxification enzymes. By complementary approaches, we show that the difference observed for the level of the porin and flagellin mRNAs in the two genetic backgrounds is perfectly correlated to the difference of proteins content. In the absence of the EIIBC^{Glc} protein, pores with larger diameter are formed in the outer membrane and the motility of cells is increased, which is expected to increase the capacity of the bacterium to seek and utilize various carbon sources. By the use of strains mutated in the genes encoding proteins EIIA^{Glc} and adenylate cyclase, we show that the regulation of porin and flagellin expression by EIIBC^{Glc} depends upon the CRP-cAMP complex and the activity of the EIIA^{Glc} protein.

Introduction.

Escherichia coli uses glucose as a preferential source of carbon in the presence of a variety of carbon sources. This favored use of glucose is dependent on its specific transport in the cells that is achieved by the glucose specific phosphoenolpyruvate-dependent phosphotransferase system (PTS) (for review : [1,2]). In this system, the EIIBC^{Glc} protein encoded by the *ptsG* gene plays a crucial role as it is responsible for the specific transport and phosphorylation of glucose. In addition to this direct role, the EIIBC^{Glc} protein is indirectly involved in regulation of the expression of several genes, encoding proteins implicated in glucose transport (*ptsG* [3,4], *ptsHI* [5]) and metabolism (*gapA* [6] and *sdhCDAB* [7]). Up to now, regulation of *ptsG* and *ptsHI* genes was found to be mediated by the Mlc protein (for review : [8]). Protein Mlc is a transcriptional repressor that binds upstream of several genes involved in glucose transport and blocks their transcription in the absence of glucose in the medium [8]. This negative regulation is relieved by the active transport of glucose through its specific PTS transporter. This regulation relieve is explained by sequestration of the Mlc negative regulator at the level of the inner membrane by the dephosphorylated EIIBC^{Glc} protein [3,4,9-11] (for review: [8]). In addition, the Mlc protein was shown to repress expression of the *manXYZ* operon, which encodes constituents of the mannose PTS that transports several hexoses, including glucose [12].

We previously found that presence of glucose in the medium enhances activity of the *gapA* P1 promoter for glyceraldehyde 3-phosphate dehydrogenase (GAPDH) production, and activity of the *gapB* P0 promoter for phosphoglycerate kinase (PGK) and fructose 1,6 bisphosphate aldolase (FBA) production ([6]; Bardey *et al.*, submitted). We previously demonstrated that expression of an active EIIBC^{Glc} protein is required to observe this activation [6]. Hence, it was important to test whether *gapA* P1 and *gapB* P0 activation was also a consequence of the sequestration of the Mlc protein by the EIIBC^{Glc} protein. Here, we show that this is not the case and this result suggests that the EIIBC^{Glc} protein could directly or indirectly participate in a mechanism of gene regulation that does not involve the Mlc protein. We hypothesized that the example of the regulation of the *gapA* and *pgk* genes

expression is not unique and that the presence or the absence of an active specific transport of glucose in *E. coli* might strongly modify gene expression in this organism. For a wide analysis of these changes of gene expression, we compared the transcriptomes of cells producing an active EIIBC^{Glc} protein and cells producing an inactive EIIBC^{Glc} protein during a growth performed in the presence of glucose.

Interestingly, we found strong variations in the expression of genes involved in pores and flagella formation. In *E. coli*, the *ompC* and *ompF* genes encode the OmpC and OmpF porins, which form relatively non specific pores that allow the rapid passage of small hydrophilic molecules across the outer membrane (for review : [13]). The difference between the two porins is that the pore formed by OmpF is wider than the pore formed by OmpC [14]. In consequence, the OmpF porin allows a much larger number of compounds to enter in the cell. The relative abundance of these porins is under an efficient regulation by environmental signals. This regulation is mainly based on a transcriptional mechanism. Indeed, the regulation of the expression of the *ompC* and *ompF* genes was found to depend upon OmpR-P, the phosphorylated form of the OmpR protein of the osmosensing two component system (for review : [15,16]). Several OmpR-P binding sites were located upstream of the *ompF* and *ompC* genes [17,18]. When the rate of OmpR-P is low, it activates transcription of the *ompF* gene by binding to the promoter region at high affinity sites [17]. When the concentration of OmpR-P increases, it binds to lower affinity binding sites situated in the *ompC* promoter region, and consequently, the transcription of the *ompC* gene is activated. Also, it binds to the *ompF* promoter region at other lower affinity binding sites and is responsible for a repression of the *ompF* transcription (for review: [16]). OmpR-P is also a negative transcriptional regulator of the flagellar master operon *flhD* and *flhC* [19] encoding the FlhD₂C₂ transcriptional activator, which controls positively the expression of the flagellar regulon in *E. coli* [20]. A first level of control concerns the *fliA* gene, which is directly and positively regulated by FlhD₂C₂. This gene encodes the flagellar σ^{28} factor, which allows transcription initiation with the E σ^{28} holoenzyme of the flagellar constituent genes, e.g. *fliC*

encoding the filament structural protein flagellin [21]. The genes of the *flhDC* operon are also negatively regulated by the H-NS protein and regulated positively by the CRP-cAMP complex [22]. This regulation is a link between motility and metabolism.

To try to understand how the observed variations of expression of porin and flagella genes are generated by the absence of an active EIIBC^{Glc} protein, we took into consideration the mechanism of transport of glucose by the glucose specific PTS. The phosphoryl group derived from the phosphoenol pyruvate is transferred sequentially to glucose by several proteins, the general PTS cytoplasmic proteins EI, HPr, and the glucose specific EIIA^{Glc} and the membrane bound EIIBC^{Glc} proteins [1,2]. EIIA^{Glc} is a central protein, which is present in the cytoplasm under two forms: a non-phosphorylated EIIA^{Glc} and the phosphorylated EIIA^{Glc}-P. The phosphorylated form is essential for catabolic repression as it regulates adenylate cyclase activity [1,2,23-25]. When the concentration of EIIA^{Glc}-P is elevated in the cytoplasm (when glucose is not present), it activates adenylate cyclase activity and increases cAMP concentration [24]. The CRP-cAMP complex, which is formed by an association of two CRP (CAP) and two cAMP, controls the expression of a large number of genes or operons. The phosphorylation state of EIIA^{Glc} is reduced when glucose is present as the phosphate is transferred to EIIBC^{Glc} and to the entering glucose. Consequently, in presence of glucose the level of CRP-cAMP is decreased. We hypothesized that the level of the phosphorylated EIIA^{Glc} protein and of the CRP-cAMP complex may be directly or indirectly involved in the observed gene regulation. To confirm this involvement, we used cells mutated in the genes *crr* and *cya* encoding respectively the EIIA^{Glc} protein and the adenylate cyclase. Here, we demonstrate that the absence of active specific transport of glucose in *E. coli* is accompanied by strong modification of the expression of several genes involved in the production of the constituents of the external membrane. We show that the absence of production of proteins EIIA^{Glc} and EIIBC^{Glc} have similar effects on the expression of these membrane components and also that the regulations observed depend upon the concentration of the CRP-cAMP complex in the cells.

Materials and Methods.

Bacterial strains and plasmids.

The *E. coli* strains JM101 (F' *traD36 lacIq* Δ (*lacZ*)M15 *proA*⁺*B*⁺/*supE* Δ (*lac-proAB*)) and IBPC1012 (JM101 *mlc*::Tc) [4] were used to determine the influence of the Mlc protein on *gapA* P1 transcription. The *E. coli* strain TP2503 (F⁻ *xyl argH1 ilvA*) and TP2504 (F⁻, *xyl argH1 ilvA zcf-229*::Tn10 *ptsG22*) [26] (gift from J. Plumbridge, Paris) were used to study the *in vivo* transcriptional activity of the different promoters. The influence of the CRP-cAMP complex was studied in strain TP2006 (F⁻ *xyl lac* Δ X74 Δ *cya glp-8306*) [27]. The role of the *ptsG* and *cya* genes products was analyzed in strain TP2512 (F⁻ *xyl argH1 ilvA zcf-229*::Tn10 *ptsG22* Δ *cya glp8306*) [26]. The role of the *crr* gene product was analyzed in strain TP2862 (F⁻, *xyl argH1 ilvA aroB* Δ (*crr*) Km^R) [28]. The pTSG4 and pTSG4(C421S) plasmids (provided by B. Erni) were used to produce the WT EIIBC^{Glc} protein or the Cys421Ser variant.

Medium and growth conditions.

The growth medium used was the minimal M63 medium [29] supplemented with thiamine (10 mg/ml) and 0.1% (wt/vol) Casamino Acids and with either glucose or pyruvate (0.4% [wt/vol]). When required, arginine was added (100 mg/ml). Ampicillin (100 mg/ml) was added for the growth of transformants. Cells were grown aerobically at 37°C, and growth was monitored by measuring the optical density at 600 nm (OD_{600nm}).

RNA isolation and analysis.

Total RNA was isolated by the hot phenol procedure [30] from cells harvested at an OD_{600nm} of 0.5. RNA extractions were always performed with the same amount of cells. Oligonucleotides were end-labelled with [γ -³²P]ATP (3000 Ci/mmol) (Amersham) and T4 polynucleotide kinase (Biolabs).

Oligonucleotide 1043 (5'-TTTCTGAGCAGCACGGAA-3'), O-fliA (5'-

GCCACAGCGAGTGTTTATCC-3'), O-*fliC* (5'-GAGTGATCAGCGAGAGGCTG-3'), O-*ompC* (5'-GAGCTGGGACCAGGAGGGAC-3') and O-*ompF* (5'GCAGGGACGATCACTGCCAG 3') were used as primers for *gapA*, *fliA*, *fliC*, *ompC* and *ompF* cDNA synthesis with reverse transcriptase. For annealing, 5 ng of end-labelled oligonucleotide and 20 µg of total RNA were heated to 65°C for 10 min in 10 µl of 50 mM Tris hydrochloride (pH 8.3), 40 mM KCl, 6 mM MgCl₂ and then cooled slowly to room temperature (10 min). The cDNA synthesis was performed as previously described [31], and cDNAs were fractionated on a 7% polyacrylamide–8 M urea sequencing gel.

Transcriptome analysis.

DNA macroarray analysis was carried out using PanoramaTM *E. coli* gene array system (Sigma Genosys). cDNAs were synthesised following the manufacturer recommendations. Total RNA (40 µg) were hybridized to hexameric random primer (Invitrogen). cDNAs were synthesised by the reverse transcriptase Superscript II (Invitrogen) in the presence of [α -³³P] dCTP, a mixture of dNTP (10 mM each) and DTT (100 mM) in SuperscriptII buffer. The reaction was carried out for 1 h at 42°C. cDNAs were purified with the Quiagen removal nucleotide kit.

After 5 min incubation of the arrays in 50 ml of SSPE buffer (10 mM sodium phosphate, pH 7.7, 0.18 M NaCl, 1 mM EDTA), prehybridization was carried out in 5 ml of hybridization solution (1x Denhardt's reagent, 5xSSPE, 2% SDS, 100 µg/ml denatured Salmon sperm DNA) for at least 1h at 65°C. Subsequently, hybridization was carried out for 12-20 h at 65°C in 5 ml of hybridization solution containing the cDNA target solution, which was denatured at 95°C for 5 min. Arrays were washed twice with 50 ml of wash buffer (0.5x SSPE, 0.2% SDS) (5 min at room temperature) and twice with 100 ml of wash buffer (20 min at 65°C). Finally, arrays were air-dried for 5 min and exposed to phosphorimager screens, which were scanned with a phosphorimager (PhosphorImager SI, Molecular Dynamics). Hybridization signal intensities were quantified with the ArrayVision software (Imaging Research Inc. Version 6.0). The data were exported in Microsoft Excel and the pixel values

for the duplicate spots were averaged. For normalisation and background correction, the overall-spot normalisation function of ArrayVision was used. Each analysis was carried out three times using two independently isolated RNA preparations and two different array couples. Genes were classified as upregulated or downregulated in the *ptsG*⁻ strain as compared to the WT strain when their difference of signal to noise ratios between the two strains were larger than 2.

Motility assays.

E. coli strains were grown overnight at 37°C in M63 medium supplemented with glucose. A drop of the same amount of cells was spotted on a semi-solid agar plate (0.35% agar in M63 medium supplemented with glucose). The plates were kept at 30°C for 12 h before measurement of the colony diameters.

Outer membrane fractions and electrophoresis.

E. coli strains were grown to an OD_{600nm} of 0.5 at 37°C in M63 medium supplemented with glucose or pyruvate (11 mM each). Cells (5x10⁸ cells) were harvested and the pellet was washed twice with 10 ml of HEPES buffer 10 mM (pH 7.4). The pellet was dissolved in the same buffer and sonicated. The disrupted membranes were pelleted by centrifugation at 32000 g for 1 h and resuspended in 200 µl of a protein denaturing buffer (Tris-HCl 80 mM pH 6.8, SDS 3%, DTT 1 mM, Glycerol 15 %, Bromophenol blue 0.005 %). Samples were heated 5 min at 95°C and centrifuged 5 min at 12000 rpm. Proteins were separated on a 12 % acrylamide urea 8 M gel. Bands were stained with Coomassie blue and destained with 10 % acetic acid and 20 % ethanol. Band intensities were quantified with Genetools software (Syngene).

Western Blot analysis.

The phosphorylation state of enzyme EIIA^{Glc} was determined with a western blotting experiment with rabbit anti-EIIA^{Glc} polyclonal antibodies (provided by E. Krin, Institut Pasteur, France) as described in [25]. *E. coli* cells were grown to an OD_{600nm} value of 0.5. Culture samples of 300 µl were taken and 20 µl of 10 N NaOH was added immediately. Samples were mixed for 10 s and 180 µl of 3 M sodium acetate (pH 5.2) and 1 ml of ethanol were added. Samples were placed at -80°C for 10 min and centrifuged at 12,000 rpm for 10 min at 4° C. The pellet was rinsed with 70% ethanol, dried at room temperature and dissolved in 100 µl of protein buffer (Tris-HCl 80mM pH 6.8, SDS 3%, DTT 1 mM, Glycerol 15 %, Bromophenol blue 0.005 %). 20 µl of the sample were separated on 15% acrylamide gel. Proteins were transferred on a nitrocellulose membrane (Biorad). Anti-EIIA^{Glc} rabbit antibodies were added in PBS-Tween buffer for 1 h and the membrane was rinsed twice with PBS-Tween buffer supplemented with the blocking reagent (Amersham Biosciences). The anti rabbit antibody which has a peroxydase activity was added to the membrane for 1 h. ECL plus kit (Amersham Biosciences) was used to reveal peroxydase activity. Band intensities were measured with Genetools software (Syngene).

Results.

Mlc does not regulate the expression of the *gapA* gene.

To test whether glucose stimulation of the transcription initiation at the *gapA* P1 promoter was dependent on the presence of the Mlc protein, we compared the *gapA* mRNAs level in a *mlc*⁺ and a *mlc*⁻ strain by primer extension analysis (Fig. 1). Transcription of the *ptsG* gene was monitored in parallel in the two genetic backgrounds as a control of the Mlc regulation. The absence of Mlc production had the previously described effect on *ptsG* expression [4]. Indeed, when glucose was not present in the medium, the level of mRNAs initiated at the *ptsG* P1 promoter was lower when the Mlc protein was produced and higher in the *mlc*⁻ background. In the presence of glucose, the repression by

Mlc of the *ptsG* transcription was relieved. In contrast, the *mlc* mutation had a very limited effect on the level of *gapA* P1 mRNAs, as well in the presence as in the absence of glucose. Only a very slight increase of the *gapA* P1 mRNA level was observed in the presence of glucose in the medium in the *mlc*⁻ strain. We concluded that the glucose effect observed for the *gapA* P1 promoter is not mediated by the Mlc protein.

Transcriptome analysis of a *ptsG*⁺ and a *ptsG*⁻ strain.

For a wider study of the effect of the absence of the EIIBC^{Glc} in *E. coli*, we carried out macroarray experiments by using the Panorama gene arrays (Sigma Genosys), which contain most of the genomic ORFs of *E. coli*. We compared the transcriptomes of the *E. coli* wild type TP2503 *ptsG*⁺ strain and the TP2504 *ptsG*⁻ strain, for which the EIIBC^{Glc} protein is not produced. The two strains were grown in M63 medium containing 0.2 % glucose until the midlog phase of growth. Total RNA was extracted and the cDNAs were synthesized and labelled with ³³P by primer extension as described in Materials and Methods. After hybridization of the arrays with the labelled cDNAs and analysis of the data, the ratio of gene expression between the *ptsG*⁺ and *ptsG*⁻ strains was calculated for every gene. Transcription of a gene was considered to be influenced by the *ptsG* gene product EIIBC^{Glc} when in at least two independent experiments the difference in the level of the mRNAs measured in the two strains was greater than 2.

The results are presented in table 1. The expression of genes encoding three classes of protein was found to be affected by the *ptsG*⁻ mutation. The first class of protein corresponds to membrane constituents, e.g. proteins involved in flagella formation and flagella constituents (*fliC*, *fliD*, *fliA* and *flgE*), porins (*ompF* and *ompC*), transport proteins (*viaO* and *putP*) and proteins associated to the membrane (*cirA*, *ompT* and *ompX*). The second class of protein are enzymes implicated in the catabolism of carbon sources (*gatY*, *gatZ*, *gltA*, *lyxK*, *kduI*, *kduD* and *thrL*), and the third class contains detoxification enzymes (*sodA* and *ahpC*). Among the membrane constituents, the *ompF* and *fliC* genes

showed the strongest increase of transcripts level upon *ptsG* mutation (Table 1). The increase of transcripts rate was of 73.3 and 15.5, respectively. Conversely, the rate of *ompC* transcripts was decreased by a factor of 2.2 in the *ptsG*⁻ strain. Among the genes encoding enzymes involved in the catabolism of carbon compounds, the *kduI*, *kduD* and *gatY* genes were the most upregulated with an increase of transcripts rate of 71.8, 6.2 and 8.2 upon *ptsG* mutation, respectively.

The *ompF*, *ompC*, *fliA* and *fliC* mRNA variations are also found by primer extension analysis.

As the expression of the *ompF*, *ompC*, *fliA* and *fliC* genes were subjected to the strongest variations, we choose these genes for further analysis of the observed regulation. First, we confirmed the effect of the *ptsG* mutation on these genes by measurement of the amounts of *ompF*, *ompC*, *fliA* and *fliC* transcripts by primer extension analysis with the reverse transcriptase, using total RNA extracted from the WT and *ptsG*⁻ strains as the template. The results are shown in figure 2. In accord with data of the macroarrays, larger amounts of *fliA*, *fliC* and *ompF* cDNAs were obtained for total RNA extracted from the *ptsG*⁻ strain. In contrast, for this total RNA fraction, a lower amount of *ompC* cDNAs was obtained. Some of the faint elongation stops observed in the lanes of the autoradiogramm displayed in Fig. 2, may be due to cleavages occurring in the mRNA *in vivo*. However, the strong signals observed were in accord with the expected transcription initiation sites defined in the Ecocyc database (<http://ecocyc.org/>) [32].

The *ptsG*⁻ strain has an outer membrane with a different protein constitution.

Next, we analysed whether the differences between the amounts of transcript that we detected by macroarray and primer extension analyses were also observed at the protein level. To test whether the OmpC and OmpF porins were differently expressed in the two strains, we used an outer membrane protein extraction technique ([33], for detail see Materials and Methods). The results obtained after fractionation of the membrane proteins by gel electrophoreses are shown in the figure 3A. As found by

quantification of the intensity of the gel bands, when glucose is transported in the cell by the EIIBC^{Glc} protein, the three major porins OmpC, OmpF and OmpA are produced in similar amounts (Fig. 3A, lane 1). In the *ptsG*⁻ strain, the OmpF/OmpA ratio increased by a factor of 3.8 compared to the WT strain, while the OmpC/OmpA ratio was decreased by a factor of 5.6 (Fig. 3A, compare lanes 1 and 2). When pyruvate was used as the carbon source, the OmpF/OmpA and OmpC/OmpA ratios were not modified by the *ptsG*⁻ mutation (Fig. 3A, compare lanes 3 and 4). However, the OmpF/OmpA ratio in the WT strain was increased by a factor 2.2 during growth on pyruvate compared to growth on glucose and conversely, the OmpC/OmpA ratio during growth on pyruvate was decreased by a factor of 1.7 as compared to growth on glucose. Hence, the modification of the rate of *ompC* and *ompF* transcripts in the *ptsG*⁻ strain led to a differential production of the porins OmpC and OmpF. In the *ptsG*⁻ background and to a lesser extent when pyruvate is the sole carbon source, the major porin present in the outer membrane is the OmpF porin.

As the *fliA* and *fliC* genes are parts of the flagella-synthesis regulon and as the amounts of *fliA* and *fliC* transcripts were upregulated in the *ptsG*⁻ background, we used a motility assay to estimate the level of flagella synthesis (Fig. 3B) [34]. Low agar plates containing glucose as the sole carbon source were used to test for cell motility (see Materials and Methods). The WT cells remained at proximity of the inoculation site whereas the *ptsG*⁻ cells spread over a larger surface of the plate. Thus, the *ptsG*⁻ strain is more motile than the *ptsG*⁺ strain in low agar plates.

Porin and flagella expressions are affected by glucose transport through its specific PTS.

The *ptsG22* mutation of the TP2504 strain (*ptsG*⁻ strain, Materials and Methods) conducts to an absence of glucose intake. To our knowledge, this apparent phenotype is explained by the non-production of the protein EIIBC^{Glc} protein in the inner-membrane. Phosphorylation of protein EIIBC^{Glc} by the PTS system occurs on the Cys421 residue [35]. Replacement of this residue by a serine conducts to the production in the inner membrane of a non-active EIIBC^{Glc} protein. Hence in

this genetic background, glucose transport and phosphorylation are abolished even though the protein is present in the inner membrane [36]. To test whether the genes regulations that we observed were dependent on the ability of the EIIBC^{Glc} to be phosphorylated, we transformed the *ptsG*⁺ and *ptsG*⁻ cells with the pTSG4 and pTSG4(C421S) plasmids, which contain the WT and mutated (Cys421Ser) *ptsG* genes, respectively. The porin expression and the motility of the transformed strains were tested. As shown in figure 4A (lanes 1 and 4), when the cells were transformed with an empty plasmid, the porin expression profile was the same as that shown in figure 3 A (lanes 1 and 2). For the *ptsG*⁺ strain, expression of the WT or mutated *ptsG* gene had no significant effect on porins production (Fig. 4A, compare lanes 2 and 3 with lane 1). In contrast, in the *ptsG*⁻ strain, whereas the expression of the WT EIIBC^{Glc} restored the WT profile of porin expression (Fig. 4A, compare lanes 4 and 5), expression of the C421S EIIBC^{Glc} protein only restored a wild type level of expression for the OmpC and OmpA porins. The OmpF porin level remained high in the *ptsG*⁻ strain transformed with plasmid pTSG4(C421S). These data show that the transport and phosphorylation of glucose influence porin expression. In particular, an active EIIBC^{Glc} protein is required to get the WT level of protein OmpF. As shown in figure 4B, the motility of a cell expressing the EIIBC^{Glc} C421S mutant is intermediate between that of the WT strain and the one of the *ptsG*⁻ strain, indicating that flagella expression is also dependent on glucose transport and phosphorylation.

A consequence of glucose transport is the dephosphorylation of protein EIIA^{Glc}, whose phosphorylated form activates the cAMP synthesis. Hence, as a result of glucose transport, the intracellular CRP-cAMP complex concentration is decreased [2]. The *flhD* and *flhC* genes encoding the transcriptional activator of flagellar genes and the *ompB* gene encoding protein OmpR that regulates the porin gene expression, are regulated by the CRP-cAMP complex (see Discussion). Hence, one likely explanation of the observed changes of the *fliA* and *fliC*, and *ompF* and *ompC* genes expression in the *ptsG*⁻ strain might be the modification of the CRP-cAMP concentration through variation of the EIIA^{Glc} phosphorylation level. To test this hypothesis, we compared the levels of porin expressions in strains

ptsG⁺ or *ptsG*⁻ mutated either in the *crr* gene coding the EIIA^{Glc} protein, or in the *cya* gene encoding adenylate cyclase (Fig. 4C). In the presence of pyruvate, the porin expression profile of the *cya*⁻ cells is different from those of both the *ptsG*⁺ and *ptsG*⁻ strains (Fig. 4C, compare lane 6 with lanes 1 and 2). This suggests the implication of the synthesis of cAMP in the observed changes of porin expression. Similar data were obtained for the *cya*⁻*ptsG*⁻ double mutant (Fig. 4C, lane 7). Indeed, glucose transport was affected in this strain, but in presence of glucose, the porin expression profile was similar to that of the *ptsG*⁺ strain (compare lanes 7 and 4 in Fig. 4C). We concluded that the CRP-cAMP complex is implicated in the difference of porin expression between the WT and *ptsG*⁻ strain.

The amount of the phosphorylated form of the EIIA^{Glc} protein was tested by a western blot analysis, using a polyclonal anti-EIIA^{Glc} antibody [25] (Fig. 4 B). In the *crr*⁻ strain, no EIIA^{Glc} was detected (Fig. 4D, lane 1). For the WT strain, in the presence of glucose, the EIIA^{Glc} protein was mainly present in its unphosphorylated form (ratio EIIA^{Glc}/EIIA^{Glc}-P = 4.3). In the *ptsG*⁻ background, the EIIA^{Glc}/EIIA^{Glc}-P ratio was of 0.65 indicating that EIIA^{Glc} was mostly phosphorylated. The EIIA^{Glc}/EIIA^{Glc}-P ratio was of 0.7 and 0.35 for the *cya*⁻ and *cya*⁻*ptsG*⁻ strains, respectively. Based on the data, we concluded that the *ptsG*⁻ and *cya*⁻ mutations increased the phosphorylation rate of the EIIA^{Glc} protein.

Discussion

Here, we show that expression of an active EIIBC^{Glc} protein strongly influences a wide series of genes in *E. coli*, and this through different mechanisms.

First of all, although the level of *gapA* P1 and *gapB* P0 transcripts are highly dependent upon the presence of glucose (Fig. 1, compare lane 1 and 3) and the increase observed in the presence of glucose depends upon the expression of an active EIIBC^{Glc} protein [6], we show that the Mlc transcription repressor is not involved in the mechanism of *gapA* P1 regulation. This is in contrast to the situation found for the glucose-regulated *ptsG* and *ptsHI* genes [4,8,37]. This suggests that genes

encoding enzymes involved in glucose metabolism and genes encoding proteins needed for glucose transport are activated in the presence of glucose by two distinct mechanisms that both depends upon the expression of an active EIIBC^{Glc} protein. In addition, by the use of DNA macroarrays we found a strong dependence of the expression of both the porin genes and the flagellin genes on the presence of an active EIIBC^{Glc} protein. These data point out a strong link between the glucose specific PTS and the outer membrane composition and architecture.

The absence of active specific transport of glucose increases cell motility.

Our transcriptome analysis reveals a significant increase in the absence of active EIIBC^{Glc} protein of the expression of the *fliE*, *fliD*, *fliA* and especially *fliC* genes, which are involved in flagella synthesis. The *fliA* gene encodes the σ^{28} factor [38] involved in transcription of the class III genes (*fliE*, *fliD* and *fliC*) that encodes components of flagella. In particular, the *fliC* gene, whose expression is the more strongly increased in the *ptsG* strain, encodes flagellin, the major constituent of flagella. Our observation of a concomitant increase of the mRNA encoding the flagella specific σ factor and of the mRNA transcribed in the presence of this factor is coherent and the data were confirmed by primer extension analysis. Furthermore, as expected, an increase activity of the genes involved in flagella synthesis was found to be accompanied by an increased motility of the cells. One can imagine that in the absence of an active specific transport of glucose, bacteria have to move rapidly to find other available carbon sources in the medium.

Which kind of regulation can explain the observed increase of flagella synthesis? The *fliA* gene encoding σ^{28} is activated by the FlhD₂C₂ transcription activator [20]. The transcription of the *flhDC* operon, is known to be activated by the CRP-cAMP complex [22]. In the absence of active transport of glucose, the level of phosphorylated EIIA^{Glc} protein in the cells is expected to increase (Fig. 4D). As the phosphorylated form of this protein activates the adenylate cyclase [1,2,23-25], the concentration of the CRP-cAMP complex in the cells is also expected to increase in these conditions. Thus,

transcription of the *flhDC* operon should be increased and consequently the *fliA* expression and *fliC*, *fliD*, *fliE* expressions are also expected to be increased in accord with the present data (Fig. 5).

The efficient displacement of the *ptsG*⁻ cells that we observed on agar plates could be due to the fact that these cells have more numerous flagella than the *ptsG*⁺ cells. In addition, the activity of proteins involved in the control of the flagellar motor could be modified in these cells compared to the WT cells. Rotation of the flagellum, i.e. the activity of the flagellar motor in *E. coli*, is controlled by protein CheY and its kinase CheA (for review : [39-42]). Protein MCP (Methyl-accepting protein) is the external sensor of diverse molecules (attractants and repellents). Proteins CheA and CheW are associated to the MCP [43]. CheA autophosphorylates and transmits the phosphoryl group on an Asp residue of CheY [44]. When CheY is phosphorylated, it dissociates rapidly from the complex and binds to the flagellar motor. The result of the interaction is a clockwise rotation of the flagella [45] and the cell is said to tumble. When an attractant binds to the MCP, the kinase activity of CheA is inhibited, CheY is not phosphorylated and the result is a counterclockwise rotation of the flagella. In this case, the cell is running forward. The global movement of the cell is an addition of periods of runs and tumbles depending on the phosphorylation state of CheA and CheY [41]. Cells that can only run tend to get trapped in the agar making small swarm, whereas cells that can tumble and run, swarm longer [34]. As shown by data of figure 3B, the *ptsG*⁻ strain makes longer swim than the WT strain. In this case, the WT cells may run forward (CCW rotation of the flagella) most of the time and can be trapped in the agar. Another possibility is that in glucose containing medium, *E. coli* cells are not flagellated [45,46]. On the other hand, the *ptsG*⁻ cells tumble more often compared to the WT cells and are not trapped in the agar. The result is that *ptsG*⁻ cells have shorter period of running and longer period of tumbling.

Another link between the *ptsG* mutation (glucose metabolism) and the cell motility may be the fact that the EI protein of the PTS system is a repressor of CheA kinase activity of MCP complexes [48]. EI protein is dephosphorylated when the carbohydrate is transported, so the activity of CheA is

inhibited and the bacterium run forward (CCW rotation of the flagella). In the case of the *ptsG*⁻ strain, EI is expected to be phosphorylated, thus CheA should autophosphorylate and the periods of CCW are expected to be shorter. The cells are likely tumbling more often and thus can escape trapping in agar, which may explain the observed longer swim of the bacteria.

In conclusion, two parameters, i.e. flagella synthesis and a different control of the flagellar motor, may act synergistically to reinforce motility of the *ptsG*⁻ cells.

The absence of active specific transport of glucose changes the porin composition of the outer membrane.

The size of the pores in the outer membrane is increased as a consequence of the absence of active glucose transport. Both by mRNA analysis and by porins extraction, we showed a strong variation of the production of the two OmpC and OmpF porins in the *ptsG*⁻ cells (Fig. 2 and 3). In contrast, the level of the OmpA porin was similar to that found in the WT cells. As the diameter of the pores formed by the OmpF proteins is quite larger than the one of pores formed by the OmpC proteins, the increased expression of OmpF in the *ptsG*⁻ cells is expected to generate large-size pores. This is another adaptation of the bacteria for capture of carbon sources other than glucose, and this when the transport of this sugar becomes inefficient. How this change in porins production can be explained?

The main regulator of the porin gene expression is the OmpR protein in its phosphorylated form. At low concentration of OmpR-P, the transcription of the *ompF* gene is activated whereas at higher concentration, the transcription of *ompF* is inhibited and the transcription of *ompC* is activated [16]. In the *ptsG*⁻ strain, as OmpF is the major porin in the outer membrane, the concentration of OmpR-P must be low (Fig. 5). The *ompB* gene that produces the OmpR protein and its kinase/phosphatase EnvZ (Fig. 5 (2)) is subjected to a complex regulation that depends on the CRP-cAMP complex [49]. The expected increase of the CRP-cAMP complex concentration in the *ptsG*⁻ cells probably modifies

the rate of production and consequently, of OmpR protein phosphorylation, which may explain the observed difference of the OmpC and OmpF genes expression.

Accordingly, when in the *ptsG*⁻ context, the adenylate cyclase is not expressed (Fig. 4C), the overexpression of the OmpF protein is not observed, showing that cAMP synthesis is required for the increased expression of OmpF in the *ptsG*⁻ cells. A second confirmation of the involvement of cAMP synthesis in the observed increased expression of OmpF is the disappearance of this increase in the absence of EIIA^{Glc} protein production, required for activation of the adenylate cyclase (Fig. 4C). Hence, altogether our data indicate that the absence of active transport of glucose through its specific PTS results in an increased of the CRP-cAMP concentration that modifies the porins genes expression. This modification reinforces the ability of cells to capture larger carbon source molecules.

Change in CRP-cAMP complex concentration may explain the change in expression of several other genes in the *ptsG*⁻ cells.

The other genes that we found to be upregulated in the *ptsG*⁻ strain like *yiaO*, *lyxK* and *cirA* may also be under the control of the CRP-cAMP complex. Indeed, the *yiaO* and *lyxK* genes belong to the same operon, the *yiaKLMNOPQRS* operon, which is probably implicated in carbohydrate metabolism [50]. The transcription of this operon is positively and directly regulated by the CRP-cAMP complex [50]. This complex is also involved in the positive regulation of transcription of the *cirA* gene whose expression was found to be increased but the mechanism is not well understood [51].

Recently, a DNA microarray approach, using the Affymetrix technology was used to identify the *E. coli* genes that are regulated by the presence of glucose in the LB medium and controlled by the CRP-cAMP complex [52]. Many operons encoding enzymes of the central metabolic pathway (Krebs enzymes), transporter and enzymes that initiate carbon metabolism appeared to be controlled by the CRP-cAMP complex. From the genes already known to be glucose-regulated only genes *ptsG* was found in this study to be subject to CRP-dependent glucose activation and the *sdhAD* genes were

found to be subject to CRP-dependent glucose repression. Implication of CRP for the regulation of the *sdhAD* genes is in contradiction with the study of [9]. Neither the *gapA*, *gapBpgkfbaA*, *ptsHI* genes nor the *fliA*, *fliC*, *ompF*, *ompC* genes were found as regulated by glucose and CRP. Utilization of different *E. coli* strains between the studies, cultures performed in rich vs minimal medium and analysis by micro vs macro arrays could explain the discrepancies of the data between the two studies.

Cross talk between two-component systems could also explain explain the change in expression of several genes in the *ptsG* cells.

Another possibility to explain the regulations observed during this work, is the potential role of other response regulators by cross talk between different two-component systems. Indeed, we found the *sodA* and *gltA* genes among the genes that are upregulated in the *ptsG*⁻ strain. Transcription of these genes is regulated by the ArcA/ArcB two-component system [52,53]. Interestingly, phosphorylation state of the OmpR/EnvZ proteins was suggested be linked to the ArcA/ArcB system [55]. Another interesting point is the fact that the *gltA* gene is adjacent to the *sdhCDAB* gene and oriented in the opposite direction. In the intergenic region, the ArcA protein has four binding sites and regulates the expression of the two divergent genes [53,55,56]. Moreover, the expression of the *sdhCDAB* operon is subject to a glucose repression that is mediated by the EIIBC^{Glc} protein [7] but the mechanism of this regulation is not known. Both promoters could be regulated by the same mechanism and the transcription may be increased in the case of the *ptsG*⁻ mutation. However, the *sdhCDAB* genes were not upregulated in the *ptsG*⁻ strain in our macroarray approach but this may be due to the rapid degradation of the mRNAs [58].

Aknowledgements

We are very grateful to B. Erni for providing us with plasmid pTSG4 and pTSG4(C421S). B. Thouvenot was a fellow of the French Ministère délégué à la Recherche et aux Nouvelles Technologies. This work was financially supported by the 'Le transcriptome d'*Escherichia coli* :

analyse globale d'une bactérie modèle' of the Centre National de la Recherche Scientifique (CNRS) and the French Ministère délégué à la Recherche et aux Nouvelles Technologies and by funds of the Region Lorraine (PRST Bioingénierie).

References

- 1 Postma, P. W., Lengeler, J. W. and Jacobson, G. R. (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**, 543-594.
- 2 Postma, P. W., Lengeler, J. W. and Jacobson, G. R. (1996) Phosphoenolpyruvate:carbohydrate phosphotransferase systems. In *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edn (Neidhart, F., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B. and Magasanik, B., eds.), pp. 1149-1174, American Society for microbiology press, Washington DC.
- 3 Kimata, K., Inada, T., Tagami, H. and Aiba, H. (1998) A global repressor (Mlc) is involved in glucose induction of the *ptsG* gene encoding major glucose transporter in *Escherichia coli*. *Mol. Microbiol.* **29**, 1509-1519.
- 4 Plumbridge, J. (1998b) Expression of *ptsG*, the gene for the major glucose PTS transporter in *Escherichia coli*, is repressed by Mlc and induced by growth on glucose. *Mol. Microbiol.* **29**, 1053-1063.
- 5 De Reuse, H. and Danchin, A. (1991) Positive regulation of the *pts* operon of *Escherichia coli*: genetic evidence for a signal transduction mechanism. *J. Bacteriol.* **173**, 727-733.
- 6 Charpentier, B., Bardey, V., Robas, N. and Branlant, C. (1998) The EII^{Glc} protein is involved in glucose-mediated activation of *Escherichia coli* *gapA* and *gapB-pgk* transcription. *J. Bacteriol.* **180**, 6476-6483.
- 7 Takeda, S., Matsushika, A. and Mizuno, T. (1999) Repression of the gene encoding succinate dehydrogenase in response to glucose is mediated by the EIICB^{Glc} protein in *Escherichia coli*. *J Biochem (Tokyo)* **126**, 354-360.
- 8 Plumbridge, J. (2002) Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. *Curr. Opin. Microbiol.* **5**, 187-193.
- 9 Lee, S. J., Boos, W., Bouche, J. P. and Plumbridge, J. (2000) Signal transduction between a membrane-bound transporter, PtsG, and a soluble transcription factor, Mlc, of *Escherichia coli*. *EMBO J.* **19**, 5353-5361.
- 10 Tanaka, Y., Kimata, K. and Aiba, H. (2000) A novel regulatory role of glucose transporter of *Escherichia coli*: membrane sequestration of a global repressor Mlc. *EMBO J.* **19**, 5344-5352.
- 11 Nam, T. W., Cho, S. H., Shin, D., Kim, J. H., Jeong, J. Y., Lee, J. H., Roe, J. H., Peterkofsky, A., Kang, S. O., Ryu, S. and Seok, Y. J. (2001) The *Escherichia coli* glucose transporter enzyme IICB^{Glc} recruits the global repressor Mlc. *EMBO J.* **20**, 491-498.
- 12 Plumbridge, J. (1998a) Control of the expression of the *manXYZ* operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. *Mol. Microbiol.* **27**, 369-380.
- 13 Nikaido, H. (1996) Outer membrane. In *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edn (Neidhart, F., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B. and Magasanik, B., eds.), pp. 29-47, American Society for microbiology press, Washington DC
- 14 Nikaido, H. and Rosenberg, E. Y. (1983) Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins. *J. Bacteriol.* **153**, 241-252.
- 15 Forst, S. A. and Roberts, D. L. (1994) Signal transduction by the EnvZ-OmpR phosphotransfer system in bacteria. *Res. Microbiol.* **145**, 363-373.

- 16 Egger, L. A., Park, H. and Inouye, M. (1997) Signal transduction via the histidyl-aspartyl phosphorelay. *Genes Cells* **2**, 167-184.
- 17 Huang, K. J. and Igo, M. M. (1996) Identification of the bases in the *ompF* regulatory region, which interact with the transcription factor OmpR. *J. Mol. Biol.* **262**, 615-628.
- 18 Head, C. G., Tardy, A. and Kenney, L. J. (1998) Relative binding affinities of OmpR and OmpR-phosphate at the *ompF* and *ompC* regulatory sites. *J. Mol. Biol.* **281**, 857-870.
- 19 Shin, S. and Park, C. (1995) Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. *J. Bacteriol.* **177**, 4696-4702.
- 20 Liu, X. and Matsumura, P. (1994) The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons. *J. Bacteriol.* **176**, 7345-7351.
- 21 Macnab, R. M. (1996) Flagella and Motility. In *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edn (Neidhart, F., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B. and Magasanik, B., eds.), pp. 123-145, American Society for microbiology press, Washington DC
- 22 Soutourina, O., Kolb, A., Krin, E., Laurent-Winter, C., Rimsky, S., Danchin, A. and Bertin, P. (1999) Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. *J. Bacteriol.* **181**, 7500-7508.
- 23 Feucht, B. U. and Saier, M. H., Jr. (1980) Fine control of adenylate cyclase by the phosphoenolpyruvate:sugar phosphotransferase systems in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **141**, 603-610.
- 24 Reddy, P. and Kamireddi, M. (1998) Modulation of *Escherichia coli* adenyl cyclase activity by catalytic- site mutants of protein IIA^{Glc} of the phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.* **180**, 732-736.
- 25 Krin, E., Sismeiro, O., Danchin, A. and Bertin, P. N. (2002) The regulation of Enzyme IIA^{Glc} expression controls adenylate cyclase activity in *Escherichia coli*. *Microbiology* **148**, 1553-1559.
- 26 De Reuse, H. and Danchin, A. (1988) The *ptsH*, *ptsI*, and *crr* genes of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system: a complex operon with several modes of transcription. *J. Bacteriol.* **170**, 3827-3837.
- 27 Roy, A. and Danchin, A. (1982) The *cya* locus of *Escherichia coli* K12: organization and gene products. *Mol. Gen. Genet.* **188**, 465-471.
- 28 Levy, S., Zeng, G. Q. and Danchin, A. (1990) Cyclic AMP synthesis in *Escherichia coli* strains bearing known deletions in the *pts* phosphotransferase operon. *Gene* **86**, 27-33.
- 29 Miller, J. H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 30 Aiba, H., Adhya, S. and de Crombrughe, B. (1981) Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* **256**, 11905-11910.
- 31 Charpentier, B. (1994) Etude des promoteurs transcriptionnels des gènes *gapA* et *gapB*, codant pour la glyceraldéhyde-3-phosphate déshydrogénase d'*Escherichia coli*. Relation structure/fonction. Thèse Université Henry Poincaré, Nancy
- 32 Karp, P. D., Arnaud, M., Collado-Vides, J., Ingraham, J., Paulsen, I. T. and Saier, M. H., Jr. (2004) The *E. coli* EcoCyc Database: No Longer Just a Metabolic Pathway Database. *ASM News* **70**, 25-30.
- 33 Liu, X. and Ferenci, T. (1998) Regulation of porin-mediated outer membrane permeability by nutrient limitation in *Escherichia coli*. *J. Bacteriol.* **180**, 3917-3922.
- 34 Wolfe, A. J. and Berg, H. C. (1989) Migration of bacteria in semisolid agar. *Proc. Natl. Acad. Sci. U S A* **86**, 6973-6977.

- 35 Meins, M., Jenó, P., Müller, D., Richter, W. J., Rosenbusch, J. P. and Erni, B. (1993) Cysteine phosphorylation of the glucose transporter of *Escherichia coli*. *J. Biol. Chem.* **268**, 11604-11609.
- 36 Nuoffer, C., Zanolari, B. and Erni, B. (1988) Glucose permease of *Escherichia coli*. The effect of cysteine to serine mutations on the function, stability, and regulation of transport and phosphorylation. *J. Biol. Chem.* **263**, 6647-6655.
- 37 Plumbridge, J. (1999) Expression of the phosphotransferase system both mediates and is mediated by *mlc* regulation in *Escherichia coli*. *Mol. Microbiol.* **33**, 260-273.
- 38 Helmann, J. D. and Chamberlin, M. J. (1987) DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative sigma factor. *Proc. Natl. Acad. Sci. U S A* **84**, 6422-6424.
- 39 Stock, J. B. and Surette, M. G. (1996) Chemotaxis. In *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edn (Neidhart, F., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B. and Magasanik, B., eds.), pp. 1103-1129, American Society for microbiology press, Washington DC.
- 40 Manson, M. D., Armitage, J. P., Hoch, J. A. and Macnab, R. M. (1998) Bacterial locomotion and signal transduction. *J. Bacteriol.* **180**, 1009-1022.
- 41 Silversmith, R. E. and Bourret, R. B. (1999) Throwing the switch in bacterial chemotaxis. *Trends Microbiol.* **7**, 16-22.
- 42 Bourret, R. B. and Stock, A. M. (2002) Molecular information processing: lessons from bacterial chemotaxis. *J. Biol. Chem.* **277**, 9625-9628.
- 43 Gegner, J. A., Graham, D. R., Roth, A. F. and Dahlquist, F. W. (1992) Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**, 975-982.
- 44 Sanders, D. A., Gillece-Castro, B. L., Stock, A. M., Burlingame, A. L. and Koshland, D. E., Jr. (1989) Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. *J. Biol. Chem.* **264**, 21770-21778.
- 45 Welch, M., Oosawa, K., Aizawa, S. and Eisenbach, M. (1993) Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. U S A* **90**, 8787-8791.
- 46 Berg, H. C. (2003) The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19-54.
- 47 Adler, J. and Templeton, B. (1967) The effect of environmental conditions on the motility of *Escherichia coli*. *J. Gen. Microbiol.* **46**, 175-184.
- 48 Lux, R., Munasinghe, V. R., Castellano, F., Lengeler, J. W., Corrie, J. E. and Khan, S. (1999) Elucidation of a PTS-carbohydrate chemotactic signal pathway in *Escherichia coli* using a time-resolved behavioral assay. *Mol. Biol. Cell* **10**, 1133-1146.
- 49 Huang, L., Tsui, P. and Freundlich, M. (1992) Positive and negative control of *ompB* transcription in *Escherichia coli* by cyclic AMP and the cyclic AMP receptor protein. *J. Bacteriol.* **174**, 664-670.
- 50 Ibanez, E., Campos, E., Baldoma, L., Aguilar, J. and Badia, J. (2000) Regulation of expression of the *yiaKLMNOPQRS* operon for carbohydrate utilization in *Escherichia coli*: involvement of the main transcriptional factors. *J. Bacteriol.* **182**, 4617-4624.
- 51 Griggs, D. W., Kafka, K., Nau, C. D. and Konisky, J. (1990) Activation of expression of the *Escherichia coli* *cir* gene by an iron-independent regulatory mechanism involving cyclic AMP-cyclic AMP receptor protein complex. *J. Bacteriol.* **172**, 3529-3533.
- 52 Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W. A. and Saier, M. H., Jr. (2004) Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J. Bacteriol.* **186**, 3516-3524.

- 53 Tardat, B. and Touati, D. (1993) Iron and oxygen regulation of *Escherichia coli* MnSOD expression: competition between the global regulators Fur and ArcA for binding to DNA. *Mol. Microbiol.* **9**, 53-63.
- 54 Park, S. J., McCabe, J., Turna, J. and Gunsalus, R. P. (1994) Regulation of the citrate synthase (*gltA*) gene of *Escherichia coli* in response to anaerobiosis and carbon supply: role of the *arcA* gene product. *J. Bacteriol.* **176**, 5086-5092.
- 55 Matsubara, M., Kitaoka, S. I., Takeda, S. I. and Mizuno, T. (2000) Tuning of the porin expression under anaerobic growth conditions by his-to-Asp cross-phosphorelay through both the EnvZ-osmosensor and ArcB-anaerosensor in *Escherichia coli*. *Genes Cells* **5**, 555-569.
- 56 Shen, J. and Gunsalus, R. P. (1997) Role of multiple ArcA recognition sites in anaerobic regulation of succinate dehydrogenase (*sdhCDAB*) gene expression in *Escherichia coli*. *Mol. Microbiol.* **26**, 223-236.
- 57 Park, S. J., Tseng, C. P. and Gunsalus, R. P. (1995) Regulation of succinate dehydrogenase (*sdhCDAB*) operon expression in *Escherichia coli* in response to carbon supply and anaerobiosis: role of ArcA and Fnr. *Mol. Microbiol.* **15**, 473-482.
- 58 Cunningham, L. and Guest, J. R. (1998) Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. *Microbiology* **144**, 2113-2123.

Figure legends

Table 1. Compilation of the data obtained by macroarray analysis of the transcriptomes of the *ptsG*⁺ and *ptsG*⁻ strains.

Analysis of data was performed with the ArrayVision software (Imaging Research Inc. Version 6.0). Differences of expression were taken into consideration if in at least two independent experiments their difference of expression between the two strains were of a factor greater than 2. The data shown are relative changes as calculated by the Excel Software.

Figure 1. The *gapA* P1 promoter is not regulate by the Mlc protein.

Analysis of the effects of the Mlc protein on the *in vivo gapA* mRNA levels by reverse transcription. The isogenic strain *E. coli* JM101 and IPBC1012 (*mlc*⁻) were grown to an OD_{600nm} of 0.5 and the total RNA were extracted. The *gapA* mRNAs were analyzed by primer extension using oligonucleotide 1043. The cDNAs corresponding to transcripts initiated at the *gapA* P1, P2, P3, P4 and *ptsG* P1 promoters are indicated.

Figure 2. Primer extension analysis of the levels of the *ompC*, *ompF*, *fliA* and *fliC* transcripts in the *ptsG* strain compared to the WT strain.

E. coli TP2503 and TP2504 strains were grown to an OD_{600nm} of 0.5. Total RNA was extracted as described in Materials and Methods and analyzed by primer extension using the O-*fliA*, O-*fliC*, O-*ompC* and O-*ompF* oligonucleotides. The cDNAs corresponding to transcripts initiated at the different promoters of *ompF*, *ompC*, *fliA* and *fliC* genes are indicated.

Figure 3. The absence of EIIBC^{Glc} has an effect on the outer membrane composition.

A) *E. coli* TP2503 and TP2504 strains were grown to an OD_{600nm} of 0.5 in M63 glucose or pyruvate medium. The porins were extracted and purified by gel electrophoresis as described in Materials and Methods. The different porins are indicated.

B) Histogram representing the effects of the *ptsG* mutation on the OmpC and OmpF porins expression. The bands were quantified as described in Materials and Methods and the relative ratios between OmpC/OmpA (dashed bars) and OmpF/OmpA (gray bars) were compared to that measured in the WT strain grown in the glucose-containing medium. Variations of porins expression were represented as percentages of the OmpC/OmpA ratio found for the WT strain. The values correspond to the mean values of three different experiments and the estimated errors are indicated by vertical bars.

C) Motility of the WT and *ptsG* strains in low agar plates containing glucose. *E. coli* TP2503 (WT) and TP2504 (*ptsG*) strains were grown overnight in M63 glucose and a drop of culture was deposit on the surface of the low agar plate. The colony diameter was measured and the given values are the mean of three experiments.

Figure 4. The EIIA^{Glc} protein and CRP-cAMP complex are involved in the difference of porin expression in the *ptsG* strain.

A and B) *E. coli* TP2503 and TP2504 strains transformed with plasmids pBR322 (-), pTSG4 (EIIBC WT) and pTSG4(C421S) (EIIBC C421S) were grown to an OD_{600nm} of 0.5 in M63 glucose supplemented with ampicillin. The porins extraction and fractionation and the measurements of cells motility were performed as in figure 3.

C) *E. coli* TP2503, TP2504, TP2862, TP2006 and TP2512 strains were grown to an OD_{600nm} of 0.5 in M63 glucose or pyruvate before porin extraction and fractionation.

D) Western blot showing the EIIA^{Glc} protein in its two phosphorylation state. *E. coli* TP2503, TP2504, TP2862, TP2006 and TP2512 strains were grown to an OD_{600nm} of 0.5 in M63 glucose. Proteins were extracted as described in Materials and Methods. The phosphorylation state of the EIIA^{Glc} protein is indicated.

Figure 5. Model explaining the variations of gene expression observed in the *ptsG* strain.

Circles represent proteins and P represents their phosphorylated state. Genes are represented by rectangle and the sites of transcription initiation by broken arrows. Activation of transcription initiation is represented by a +.

Gene	Increase or decrease factor	Class of molecule	Identity of the encoded protein
Upregulated in the <i>ptsG</i> strain			
<i>yiaO</i>	3.0	Unknown	Homologous to transport protein
<i>putP</i>	2.1	Transport protein	Proline/sodium symporter
<i>ompF</i>	73.3	Membrane constituent	OmpF porin
<i>fliC, fliD, fliA</i>	15.5, 4.0, 2.0	Membrane constituent	Flagellar structure
<i>flgE</i>	2.4	Membrane constituent	Flagellar structure
<i>cirA</i>	5.5	Membrane constituent	Colicin receptor
<i>ompT</i>	5.3	Membrane constituent	Protease VII precursor
<i>gatY, gatZ</i>	8.2, 2.1	Catabolism of carbon compound	Tagatose biphosphate aldolase
<i>gltA</i>	3.0	Catabolism of carbon compound	Citrate synthase TCA cycle
<i>lyxK</i>	2.1	Catabolism of carbon compound	L-xylulokinase
<i>kduI</i>	71.8	Catabolism of carbon compound	Homolog to pectin degrading enzyme 5-keto 4-deoxyuronate isomerase
<i>kduD</i>	6.2	Catabolism of carbon compound	2-deoxy-D-gluconate 3- dehydrogenase
<i>sodA</i>	12.0	Detoxification	Superoxyde dismutase
<i>ahpC</i>	3.9	Detoxification	Alkylhydroperoxide reductase
Downregulated in the <i>ptsG</i> strain			
<i>ompC</i>	2.2	Membrane constituent	OmpC porin
<i>ompX</i>	2.0	Membrane constituent	Protein X precursor
<i>thrL</i>	2.2	Amino acids biosynthesis and metabolism	Operon <i>thr</i> peptide leader

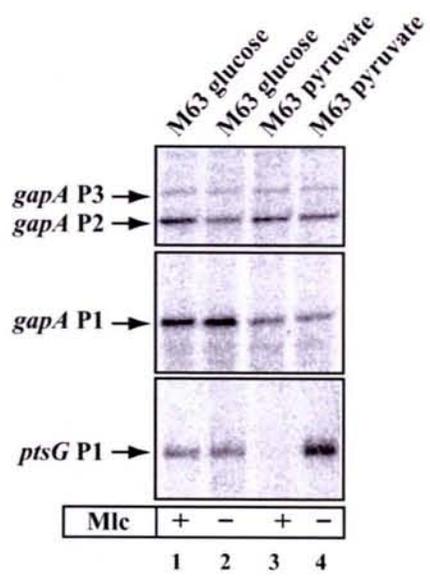


Figure 1 Thouvenot et al, 2004

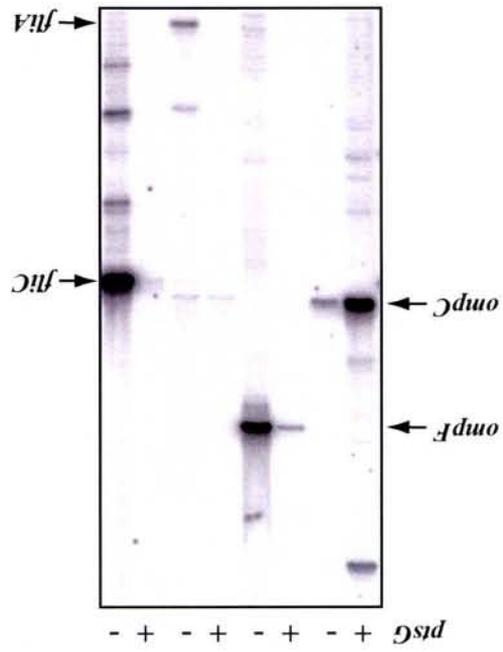


Figure 2 Thouvenot et al, 2004

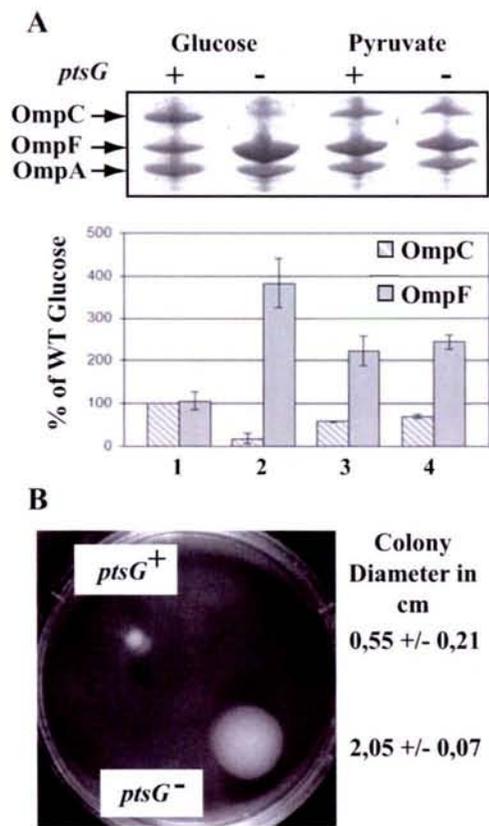


Figure 3 Thouvenot et al, 2004

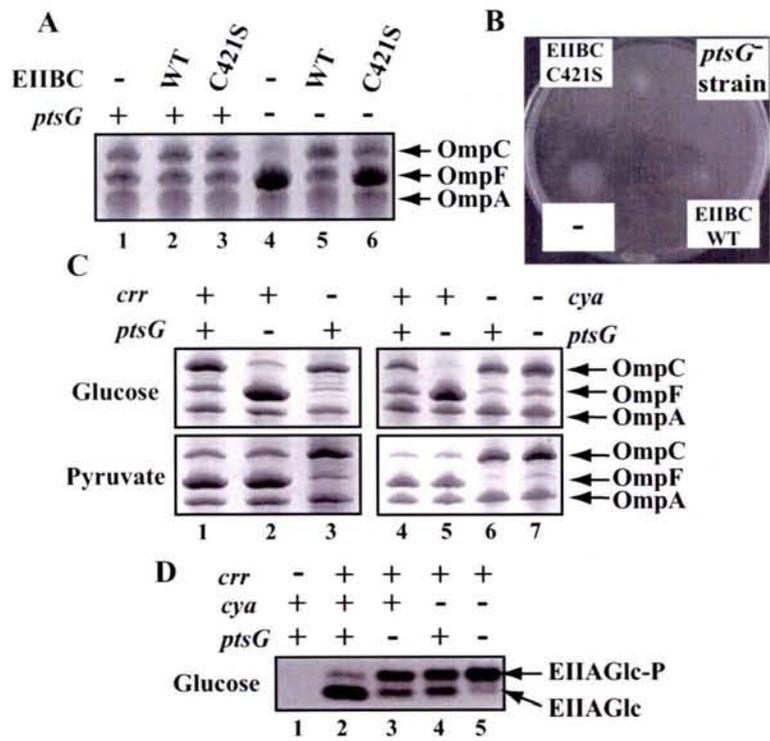


Figure 4 Thouvenot et al, 2004

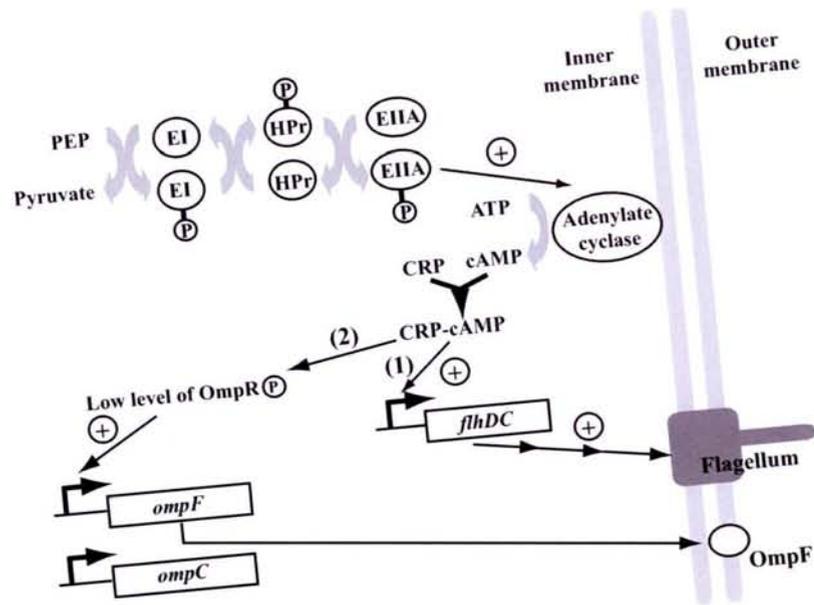


Figure 5 Thouvenot et al, 2004

III.) Discussion.

Les données que nous avons obtenues montrent qu'en absence de la protéine EIIBC^{Glc} et donc de perméation du glucose par le système PTS, l'expression de gènes codant les protéines de constitution des flagelles est augmentée, ce qui rend la bactérie plus mobile. D'autre part, le profil d'expression des porines est modifié et la porine OmpF qui permet de former des pores de plus grand diamètre est majoritaire dans la membrane externe de la bactérie. Ces caractéristiques des bactéries *ptsG* en croissance dans un milieu contenant du glucose seraient similaires à celles observées pour une souche où le système PTS spécifique du glucose est actif mais qui est cultivée en absence de glucose. Les données de la littérature montrent en effet qu'en présence de glucose, les cellules ne sont pas flagellées (pour revue : Soutourina et Bertin, 2003; Berg, 2003), ce qui impliquerait qu'en absence de glucose, les bactéries seraient plus mobiles pour rechercher de nouvelles sources de carbone. De plus, le diamètre des pores formés dans la membrane est plus grand afin de permettre une entrée de nutriments de grande taille ou en quantité plus importante (Liu et Ferenci, 1998). De plus, des gènes codant des enzymes impliquées dans le catabolisme de sources de carbone autres que le glucose sont exprimés, ce qui doit permettre un changement du métabolisme de la bactérie qui s'adapte ainsi au changement de nutriments.

D'après nos données et celles de la littérature, un des éléments principaux dans le processus d'adaptation que nous avons observé dans la souche *ptsG* est la concentration en complexe CRP-AMPC. Dans le cas de la souche *ptsG*, la concentration de ce complexe doit être fortement augmentée du fait que la protéine EIIA^{Glc} doit être majoritairement présente sous forme phosphorylée. Nos données suggèrent aussi que, dans la souche *ptsG*, la protéine OmpR serait majoritairement présente sous sa forme phosphorylée.

III.1) Régulation de la mobilité bactérienne.

Comme nous l'avons écrit dans la discussion de l'article, deux facteurs pourraient expliquer l'augmentation de mobilité des bactéries *ptsG* :

- 1) ces bactéries ont un nombre de flagelles plus grand,
- 2) l'activité du moteur de chaque flagelle pourrait être augmentée.

Nous émettons l'hypothèse qu'en présence de glucose dans le milieu de culture, les cellules ne possèderaient pas de flagelles du fait de la faible concentration en complexe CRP-AMPC qui est requise pour transcrire l'opéron *flhDC* codant le facteur de transcription FlhD₂C₂. Dans le cas de la souche *ptsG⁻*, l'augmentation attendue du taux de complexe CRP-AMPC du fait du taux élevé de la protéine EIIA^{Glc}-P est supposé entraîner une augmentation de la transcription de l'opéron *flhDC* qui est régulée par ce complexe (Fig. 72). En utilisant la technique d'extension d'amorce par la transcriptase inverse et un oligonucléotide complémentaire au transcrit *flhDC* (Soutourina *et al.*, 1999), nous avons vérifié que le taux de transcrits *flhDC* P1 est supérieur dans la souche *ptsG⁻* en comparaison de celui dans la souche *ptsG⁺* lorsque ces souches sont cultivées en présence de glucose (Fig. 73). Il est à noter que les taux de transcrits initiés au promoteur *flhDC* P1 dans les souches *ptsG⁺* et *ptsG⁻* cultivées en présence de pyruvate sont similaires et intermédiaires entre ceux obtenus pour ces mêmes souches en présence de glucose. De plus, nous avons vérifié que le taux de transcrits *flhDC* dans les souches *cya⁻* et *cya⁻ ptsG⁻* est très faible, aussi bien en présence de glucose qu'en présence de pyruvate (Fig. 73). Ces résultats sont en accord avec l'idée que l'augmentation du taux de synthèse des composants des flagelles dans la souche *ptsG⁻* est liée à une augmentation du taux de facteur FlhD₂C₂ dans les cellules, augmentation liée à l'augmentation de la concentration en complexe CRP-AMPC.

Nous n'avons pas réalisé d'expérience permettant de tester spécifiquement les variations de l'activité du moteur des flagelles dans la souche *ptsG⁻* par rapport à la souche *ptsG⁺*. L'utilisation de souches d'*E. coli* exprimant des variants des protéines impliquées dans la régulation de l'activité du moteur du flagelle (CheA, CheY...) devrait nous apporter des informations dans ce domaine. En parallèle, l'utilisation d'une technique de microscopie électronique pour visualiser des bactéries colorées négativement par de l'acide phosphotungstique nous permettra de vérifier l'augmentation du nombre de flagelles.

III.2) Régulation de l'expression des porines OmpC et OmpF.

En ce qui concerne les variations d'expression des porines dans les souches *ptsG⁻*, les données de la littérature suggèrent qu'elles sont liées aux variations du taux de protéines OmpR phosphorylées. D'après les données sur la régulation des gènes des porines OmpF et OmpC, pour que OmpF soit fortement exprimé dans la souche *ptsG⁻*, il faut que le taux de OmpR-P soit faible afin que seuls les sites de plus forte affinité pour OmpR-P soient occupés au niveau des éléments régulateurs du gène *ompF* (sites F1, F2 et F3, cf. introduction).

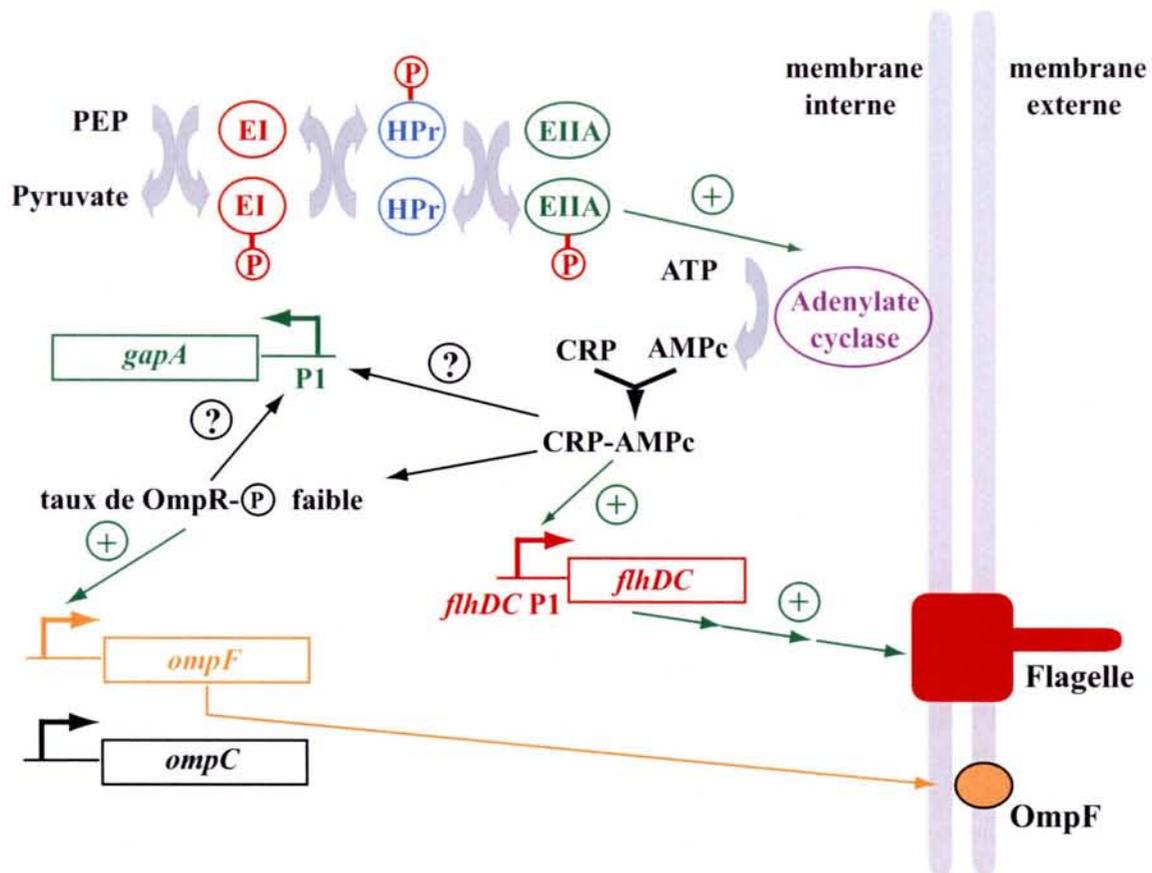


Figure 72 : Régulations pouvant expliquer les variations de l'expression des gènes *ompF*, *ompC* et des gènes impliqués dans la synthèse des flagelles dans la souche *ptsG⁻*.

Les protéines générales EI et HPr sont indiquées. Le domaine EIIA^{Glc} est représenté en vert. Le phosphate qui est transféré à partir du PEP est indiqué par un P rouge. L'adénylate cyclase est représentée en violet. Les membranes internes et externes sont représentées en gris, tandis que le flagelle et la protéine OmpF sont représentés, respectivement, en rouge et orange. Les gènes sont indiqués par des rectangles de couleurs différentes et les sites d'initiation de la transcription sont représentés par des flèches de couleurs différentes. Les régulations positives directes sont indiquées par des flèches et des "+" verts. Les régulations potentielles pouvant s'exercer sur le promoteur *gapA* P1 sont indiquées par des points d'interrogation.

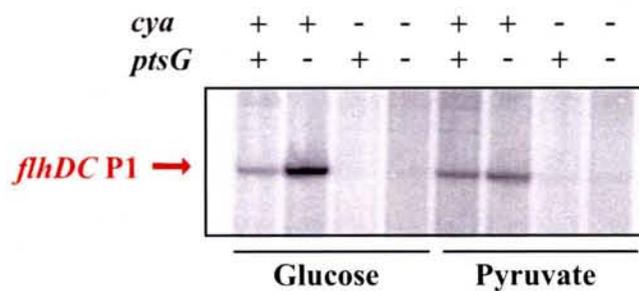


Figure 73 : Analyse par élongation d'amorce par la transcriptase inverse des variations du taux d'ARNm *flhDC P1* au cours d'une culture des souches TP2503, TP2504 (*ptsG*⁻), TP2006 (*cya*⁻) et TP2512 (*cya*⁻ *ptsG*⁻). Ces souches ont été cultivées en milieu M63 contenant, soit du glucose, soit du pyruvate comme seule source de carbone et des prélèvements de cellules ont été effectués lorsque la DO_{600nm} a atteint une valeur de 0,5. Les ARN totaux ont été extraits par la méthode au phénol chaud. Dans chaque essai, 20 µg d'ARN totaux ont été analysés par extension d'amorce par la transcriptase inverse. L'amorce utilisée est un oligonucléotide complémentaire à la séquence de l'ARNm transcrit à partir du promoteur *flhDC P1* utilisé par Soutourina *et al.* (1999) dont la séquence est 5' GGTATGCATTATTCCCACCC 3'. Les ADNc ainsi obtenus ont été fractionnés en gel de polyacrylamide dénaturant. La nature de l'ADNc est indiquée à gauche du gel.

Un grand nombre de facteurs peuvent influencer le taux de phosphorylation de la protéine OmpR (Liu et Ferenci, 2001). Le taux de phosphorylation de OmpR est en grande partie lié à l'activité de l'HK EnvZ (cf. Introduction). Comme les cultures des deux souches *ptsG*⁺ et *ptsG*⁻ ont été réalisées dans le même milieu, on peut supposer que l'activité de EnvZ qui dépend de l'osmolarité du milieu n'est pas fortement modifiée entre les deux souches. Néanmoins, la transcription de l'opéron *ompB* qui permet la production des protéines EnvZ et OmpR semble régulée directement par le complexe CRP-AMPc (Fig. 72) (Huang *et al.*, 1992) (cf. Introduction). Il existerait plusieurs promoteurs en amont du gène *ompB* qui seraient régulés de façon différentielle par le complexe CRP-AMPc (Huang *et al.*, 1992). De fortes variations de la concentration en complexe CRP-AMPc peuvent entraîner des changements de la concentration de OmpR-P et donc l'augmentation du taux de complexe CRP-AMPc dans la souche *ptsG*⁻ pourrait en partie expliquer l'effet observé sur l'expression des porines (Fig. 72). D'autres facteurs sont impliqués dans la régulation du taux de phosphorylation de OmpR. Ainsi, il a été montré que l'acétyl phosphate peut transmettre son phosphate à OmpR (Heyde *et al.*, 2000). Le pyruvate, le produit de la glycolyse (Fig. 41), est transformé en acétyl coA par la pyruvate déshydrogénase. Ce produit est ensuite soit pris en charge par le cycle de Krebs, soit transformé en acétyl phosphate (AcP) par la phosphotransacétylase codée par le gène *pta*. L'AcP est ensuite transformé en acide acétique par l'acétate kinase codée par le gène *ackA* (pour revue : McCleary *et al.*, 1993). En fonction de la source de carbone utilisée par *E. coli*, on peut s'attendre à des variations du taux d'AcP produit et donc à une variation du taux de phosphorylation de la protéine OmpR par l'AcP.

D'autres HK comme ArcB peuvent aussi phosphoryler OmpR *in vivo* (Matsubara *et al.*, 2000). Ce phénomène d'échange de phosphate entre les différents systèmes à deux composants pourrait être assez répandu, ce qui pourrait représenter un moyen de réguler finement l'expression des gènes (Matsubara *et al.*, 2000; Oshima *et al.*, 2002). Une explication possible de la diminution du taux de OmpR-P dans la souche *ptsG*⁻ pourrait être une diminution de l'activité kinase d'une autre HK.

III.3) Est-ce que les variations du taux de complexe CRP-AMPc et du taux de phosphorylation de la protéine OmpR sont responsables de l'effet du glucose sur les promoteurs *gapA* P1 et *gapB* P0 ?

L'approche puces à ADN que nous avons utilisée pour comparer les transcriptomes des souches *ptsG*⁺ et *ptsG*⁻, ne nous a pas permis de détecter de variations du taux de transcrits

initiés aux promoteurs *gapA* P1 et *gapB* P0 et nous n'avons pas détecté non plus de variations des taux de transcrits *ptsG* et *ptsHI*. Dans le cas du gène *gapA*, l'existence d'autres promoteurs transcriptionnels peut masquer les effets existants sur l'un d'eux. Dans le cas du promoteur *gapB* P0, comme nous l'avons vu, le transcrit est clivé par la RNase E puis dégradé par le dégradosome dans un grand nombre de conditions de culture, ce qui limite la détection de variations existant au niveau transcriptionnel. Le complexe CRP-AMPc est un des régulateurs transcriptionnels dont l'activité varie entre les souches *ptsG*⁻ et *ptsG*⁺. Néanmoins, les données antérieures à la fois sur le promoteur *gapA* P1 et sur le promoteur *gapB* P0 suggèrent que ce complexe n'est pas directement impliqué dans la régulation de l'activité de ces promoteurs. Ce n'est sans doute pas les variations du taux de complexe CRP-AMPc qui peuvent expliquer l'absence d'activation en présence de glucose mis en évidence antérieurement pour les promoteurs *gapA* P1 et *gapB* P0 en absence de protéine EIIBC^{Glc} (Fig. 72).

Le facteur de transcription FlhD₂C₂ dont le taux est augmenté dans la souche *ptsG*⁻ ne serait pas impliqué dans la régulation de l'activité des promoteurs *gapA* P1 et *gapB* P0 (Fig. 72). En effet, une approche par puces à ADN a été utilisée pour mettre en évidence les gènes régulés par ce facteur de transcription (Pruss *et al.*, 2001) et les auteurs n'ont pas observé de régulation des gènes *gapA*, *epd*, *pgk* ou *fbaA* en fonction de la production de ce facteur de transcription (Pruss *et al.*, 2001).

Il sera important de tester si les variations supposées du taux de protéine OmpR-P pourraient avoir une influence directe ou indirecte sur l'activité des promoteurs *gapA* P1 et *gapB* P0. La protéine OmpR-P reconnaît des séquences ADN riches en résidus T (tTTACTTTTTg-aACAT-tt) (Huang et Igo, 1996; Egger *et al.*, 1997). Nous avons recherché l'existence de ce type de séquences en amont des promoteurs *gapA* P1 et *gapB* P0 (Fig. 74). Seuls deux sites potentiels, pour lesquels l'homologie avec la séquence consensus est faible, ont été détectés en amont de ces promoteurs. Néanmoins, dans le cas du promoteur *bolA1* P régulé négativement par la protéine OmpR-P, le site de fixation de OmpR-P présente peu d'homologie avec la séquence consensus (Fig. 74) (Yamamoto *et al.*, 2000). Des expériences complémentaires devront être réalisées pour tester si OmpR-P peut se fixer en amont des promoteurs *gapA* P1 et *gapB* P0 et réguler positivement l'activité de ces promoteurs. Une possibilité sera d'utiliser des variants d'*E. coli* ne produisant pas le facteur OmpR (souches JB101 ou RN101) (Yamamoto *et al.*, 2000) ou produisant des protéines OmpR mutées soit dans le site de fixation de la protéine à l'ADN, soit dans le site de phosphorylation. L'étude

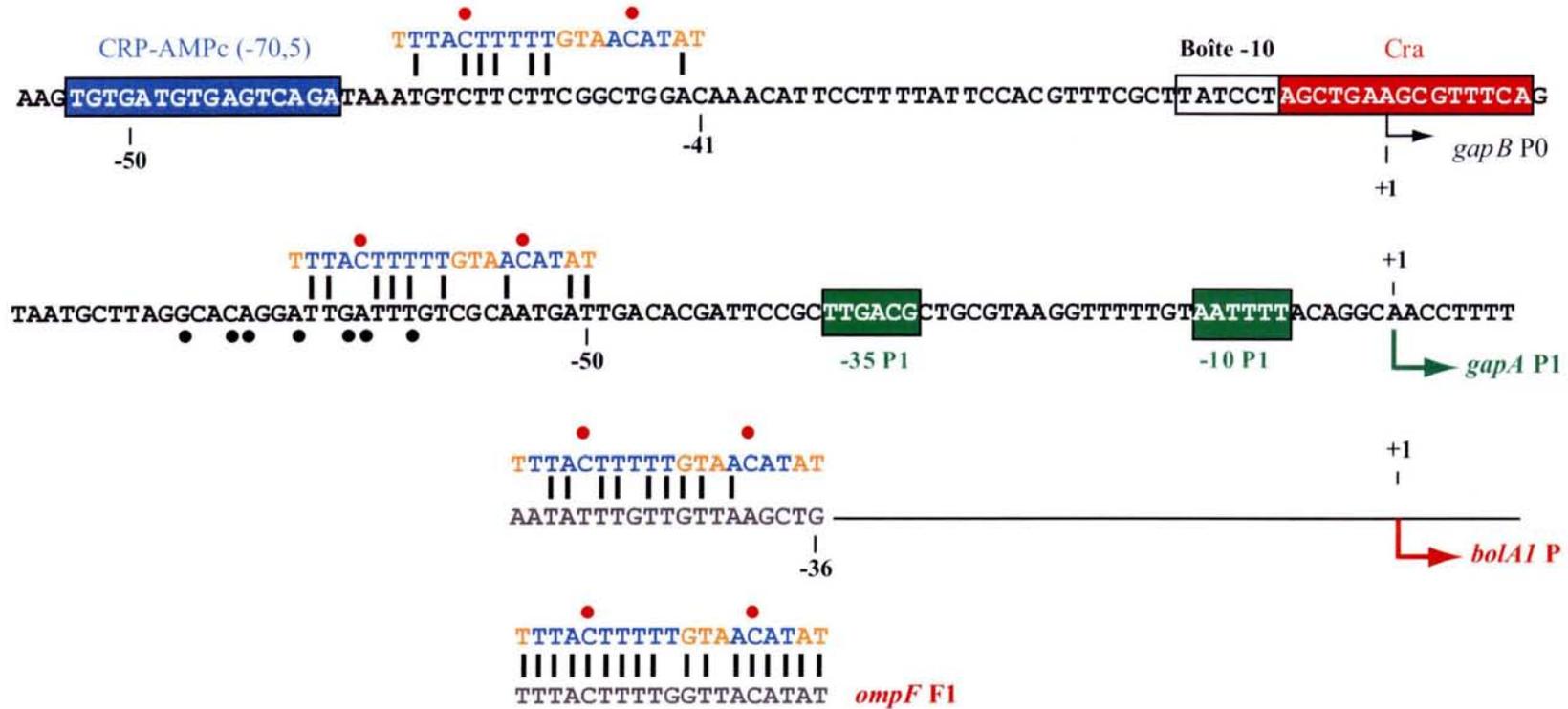


Figure 74 : Séquences et positions des sites potentiels de fixation du facteur de transcription OmpR-P en amont des promoteurs *gapA* P1 et *gapB* P0. Les sites d'initiation de la transcription sont indiqués par des flèches de couleurs différentes : *gapA* P1 en vert, *gapB* P0 en noir et *bolA1* P en rouge. Les boîtes -10 et -35 des promoteurs sont encadrées. Les sites potentiels de fixation de la protéine FIS, CRP-AMPc et Cra sont représentés, respectivement, par des points noirs, des rectangles bleus et des rectangles rouges. La séquence consensus des sites de fixation de la protéine OmpR-P est représentée au-dessus de la séquence des promoteurs. Les bases les plus conservées sont indiquées en bleu et les moins conservées en orange. Les bases identiques sont indiquées par des traits. Les points rouges représentent les résidus C conservés au niveau des sites de fixation en amont des gènes *ompF* et *ompC*. Le site de forte affinité pour OmpR-P, *ompF* F1, est indiqué.

sur l'augmentation du taux de transcrit au niveau des promoteurs *gapA* P1 et *gapB* P0 en présence de glucose permettra de savoir si la protéine OmpR est impliquée dans ces variations. Des expériences de transcription *in vitro* en présence de la protéine OmpR-P purifiée pourront être réalisées pour vérifier son implication directe dans l'initiation de la transcription aux promoteurs *gapA* P1 et *gapB* P0.

En ce qui concerne une implication possible du taux d'AcP dans l'activité des promoteurs *gapA* P1 et *gapB* P0, nous avons débuté l'étude basée sur l'emploi de variants d'*E. coli*, l'un ne produisant pas l'enzyme AckA (*ackA*⁻), ce qui conduit à une accumulation d'AcP et l'autre *ackA*⁻ *pta*⁻ ne produisant pas les deux enzymes. Dans ce cas, le taux d'AcP est faible. Ces souches nous ont été fournies par P. Laloi (Lyon). Ces variants ont été cultivés dans un milieu contenant du glucose. Les premiers résultats ne sont pas en faveur d'une implication de ces enzymes dans les régulations de l'initiation de la transcription au promoteur *gapA* P1. Il est à noter que les taux des porines OmpF et OmpC que nous avons mesurés pour ces souches varient peu. Ces données suggèrent que l'AcP ne serait pas l'élément régulateur qui permettrait l'augmentation du taux de transcrits initiés aux promoteurs *gapA* P1 et *gapB* P0.

Il a été montré récemment que la protéine ArcA-P était un répresseur spécifique de l'initiation de la transcription au promoteur *ptsG* P1 (Jeong *et al.*, 2004). L'HK qui permet de phosphoryler ArcA est la protéine ArcB. Son activité d'autophosphorylation est régulée par les quinones du système de transport des électrons dans la membrane (cf. Introduction).

Nos données de macroarray montrent une variation du taux d'expression des gènes *sodA* et *gltA* dans la souche *ptsG*⁻, ces gènes étant connus pour être régulés par ArcA (Park *et al.*, 1994; Tardat et Touati, 1993). Une hypothèse serait que le taux de phosphorylation de ArcA varie entre les souches *ptsG*⁺ et *ptsG*⁻ et que ArcA-P régule l'initiation de la transcription aux promoteurs *gapA* P1 et *gapB* P0. Cette hypothèse pourra être testée en utilisant des variants d'*E. coli* (YJ2003) (Jeong *et al.*, 2004) ne produisant plus cette protéine, et en testant le taux d'initiation de la transcription aux promoteurs *gapA* P1 et *gapB* P0 dans ces mutants.

En conclusion, cette approche par puces à ADN nous a permis de mettre en évidence les modifications de l'expression des gènes qui découlent de l'absence du transporteur spécifique du glucose chez *E. coli*. Les résultats que nous avons obtenus ne permettent pas de mettre en évidence de faibles variations de l'expression des gènes. Nous pourrions utiliser d'autres techniques de puces à ADN comme les microarray basés sur d'autres types de moyens de détection des hybridations (fluorescence) pour approfondir les données obtenues par les macroarrays.

Conclusion Générale

D'une manière générale, nos travaux ont porté sur les régulations transcriptionnelles et post-transcriptionnelles de gènes impliqués dans la glycolyse ou l'adaptation des bactéries à des variations de la disponibilité en nutriments. Nous avons d'abord réalisé une étude sur le promoteur *gapA* P1 responsable de la très forte expression du gène *gapA* en présence de glucose dans le milieu de culture.

Nous montrons que l'architecture de ce promoteur est originale du fait de la présence d'un dinucléotide TG essentiel à l'activité et d'une boîte -35, elle aussi essentielle à l'efficacité. La nécessité de ces deux séquences est liée à la présence d'une boîte -10 dont la séquence est divergente de la séquence consensus. C'est la première démonstration expérimentale de la double nécessité d'un dinucléotide TG et d'une boîte -35 pour un promoteur d'*E. coli*. En plus de ces particularités, nous avons mis en évidence des contacts à longue distance entre l'ARN pol et l'ADN du promoteur, ainsi que la présence d'une courbure intrinsèque de l'ADN située au niveau de la région interboîte qui intervient fortement dans l'efficacité du promoteur. D'après nos données, ces différentes caractéristiques structurales du promoteur *gapA* P1 permettent la formation d'un complexe ouvert RPo très stable, ce qui conduit à une forte expression du gène *gapA*. Du fait de l'originalité de ce promoteur et du grand nombre de données acquises, il sera intéressant d'approfondir l'étude de son mécanisme d'initiation de la transcription par l'analyse de la cinétique d'apparition des différents intermédiaires de formation du complexe d'initiation actif, ceci pour le promoteur sauvage et les différents variants que nous avons générés. Nous avons détecté un grand nombre de contacts entre l'ARN pol et le promoteur *gapA* P1 et il serait intéressant de préciser leur importance dans le mécanisme d'initiation en utilisant des variants de l'ARN pol dans lesquels des régions différentes des sous-unités α , β , β' ou σ sont mutées ou délétées.

L'efficacité d'initiation de la transcription au promoteur *gapA* P1 peut être améliorée en convertissant la séquence de la boîte -10 en une séquence consensus. Cela signifie qu'en dépit de son efficacité, ce promoteur n'est pas complètement optimisé, ce qui permet d'envisager l'existence de mécanismes de régulation. Ceci est en accord avec notre observation d'un rôle activateur possible de la protéine Fis, protéine fortement exprimée en début de phase exponentielle de croissance. Le facteur d'activation par Fis est faible, mais cette activation du promoteur *gapA* P1 pourrait néanmoins être importante pour optimiser la production de GAPDH en début de phase exponentielle de croissance. Par ailleurs, en dépit de nos essais multiples, nous n'avons toujours pas réussi à savoir par quel mécanisme le promoteur *gapA* P1 est activé en présence de glucose. Nous avons montré qu'il ne s'agit pas d'une levée d'inhibition par la protéine Mlc, ni par le système CsrA/CsrB.

Nous avons mis en évidence la présence d'un autre promoteur faisant partie de la famille des promoteurs étendus dans cette région inter-ORF *gapA-yeaA*. Il s'agit du promoteur *yeaA* P1 qui possède un dinucléotide TG essentiel en amont d'une boîte -10. A la différence du promoteur *gapA* P1, la boîte -10 du promoteur *yeaA* P1 a une séquence proche de celle du consensus, ce qui doit expliquer la faible importance de la boîte -35. Ce promoteur permet la transcription du gène *yeaA* codant la méthionine sulfoxyde réductase B (MsrB). Nous avons montré que l'initiation de la transcription au promoteur *yeaA* P1 est régulée négativement par la protéine Fis dont un site de fixation chevauche le site d'initiation de la transcription. Par contre, elle est régulée positivement par le complexe CRP-AMPc dont le site de fixation est centré 91 pb en amont du site d'initiation. Dans la région inter-ORF *yeaA-gapA*, un même complexe CRP-AMPc sert à la fois d'activateur de la transcription pour le promoteur *gapA* P3 et d'activateur de la transcription pour le promoteur *yeaA* P1. Il sera intéressant d'obtenir des informations sur l'organisation relative du complexe CRP-AMPc et des deux ARN pol fixées aux deux sites d'initiation de la transcription. De plus, il sera aussi intéressant de connaître le rôle de la forte courbure de l'ADN centrée au niveau du site de fixation du complexe CRP-AMPc sur l'initiation de la transcription à ces promoteurs *gapA* P3 et *yeaA* P1 en présence ou non des protéines Fis et du complexe CRP-AMPc.

L'étude de l'expression du gène *yeaA* nous a conduit à nous intéresser à la régulation de l'initiation de la traduction de son ARNm par le petit ARN non codant RyhB. Nous avons montré qu'*in vitro*, l'ARNnc RyhB interagit avec l'ARNm *yeaA* et que le site d'interaction chevauche le codon d'initiation de la traduction et le site de fixation du ribosome, ce qui suggère un blocage de la traduction par l'ARNnc RyhB. Cette hypothèse devra être vérifiée à la fois *in vitro* et *in vivo*. Il sera, en particulier, important de déterminer si l'ARNnc RyhB peut bloquer la traduction de l'ARNm *yeaA* en présence de fer dans le milieu de croissance. La protéine Hfq interagit avec l'ARNm *yeaA* mais nous n'avons pas défini précisément les sites de fixation de cette protéine. Par ailleurs, contrairement à ce qui a été observé pour beaucoup d'interactions entre ARNnc et ARNm, la protéine Hfq ne permet pas de faciliter l'interaction entre l'ARNnc RyhB et l'ARNm *yeaA*.

En ce qui concerne la co-régulation de l'expression des enzymes de la glycolyse GAPDH, PGK et FBA en présence de glucose dans le milieu de culture, nous avons montré par production de fusions transcriptionnelles *gapA* P1-*lacZ* et *gapB* P0-*lacZ* et par l'étude de la maturation des transcrits *epd-pgk-fbaA*, que cette régulation a lieu à la fois au niveau transcriptionnel et au niveau post-transcriptionnel.

La transcription des ORF *epd*, *pgk* et *fbaA* est initiée à partir du promoteur *gapB* P0 localisé en amont des 3 ORF et du promoteur *pgk* P1 situé dans l'ORF *epd*. Les transcrits initiés au promoteur *gapB* P0 sont maturés par la RNase E dans l'ORF *epd* et la région inter-ORF *epd-pgk*, ce qui entraîne la destruction des ARNm *epd* et participe à la faible expression de l'E4PDH. Nous avons montré que l'augmentation du taux de transcrits initiés aux promoteurs *gapA* P1 et *gapB* P0 en présence de glucose, qui est dépendante de la production du transporteur spécifique du glucose EIIBC^{Glc}, est à la fois un mécanisme d'ordre transcriptionnel et d'ordre post-transcriptionnel. Il nous reste maintenant à identifier la nature des mécanismes post-transcriptionnels impliqués.

L'utilisation de puces à ADN pour comparer les transcriptomes d'une souche *ptsG*⁻ et d'une souche *ptsG*⁺ nous a permis de montrer que l'absence d'expression du transporteur spécifique du glucose a des effets sans doute très similaires à ceux ayant lieu en absence de glucose. Ces données nous ont permis de formuler des hypothèses sur les mécanismes de régulation de l'expression des gènes d'*E. coli* qui permettent une adaptation de la cellule aux conditions de changements de source de carbone. Nous avons en effet montré que l'absence du transporteur spécifique du glucose dans la membrane conduit à une mobilité plus grande des cellules, mais aussi à une augmentation de la perméabilité de la membrane externe. Les hypothèses que nous avons formulées devront être vérifiées expérimentalement par emploi d'une série de variants d'*E. coli*, mutés dans les gènes dont les produits pourraient intervenir dans les régulations observées. Grâce aux résultats obtenus, il sera peut-être alors possible de comprendre par quels mécanismes l'activité des promoteurs *gapA* P1 et *gapB* P0 est augmentée en présence de glucose.

Matériels et Méthodes

Partie 1 : Matériels.

I.) Souches bactériennes utilisées.

E. coli XL1 blue : *recA1, lac, endA1, gyrA96, thi-1, hsdR17, supE44, relA-1*, (F'*proAB, lacI^q, lacZΔM15, Tn10*) (Bullock *et al.*, 1987).

E. coli TG1 : *supE, hsdD5, thi-, Δ(lac-pro AB)*, F'(traD36 *proAB+lac Iq lacZΔM15*) (Stratagène).

Ces souches ont été utilisées comme souches de référence pour toutes les analyses de la régulation de l'expression des gènes *gapA, epd* et *pgk*.

E. coli IBPC5321 : *thi-1, argE3, ΔlacX74, mlt-1i, xyl-5, tsx-29, rpsL, argG6, his-4* (Plumbridge et Soll, 1989).

E. coli IBPC6311 : *thi-1, argE3, ΔlacX74, mlt-1, xyl-5, tsx-29, rpsL, argG6, his-4, recA56, srl-300:Tn10* (Plumbridge et Soll, 1989).

Ces souches ont été utilisées pour l'obtention de phages λ recombinants et des souches monolysogènes. Elles nous ont été fournies par M. Springer (IBPC, Paris).

E. coli JM101 : *Δ(lac-proAB), thi-1, supE*, F'(traD36, *lacIq, lacZΔM15*) (Yanisch-Perron *et al.*, 1985).

E. coli JM101, *mlc::tc*

Ces souches ont été utilisées afin d'étudier l'influence de la protéine Mlc sur l'efficacité du promoteur *gapA P1*. Elles nous ont été fournies par J. Plumbridge (IBPC, Paris).

E. coli TP2503 : F' *xyl argH1 ilvA* (De Reuse et Danchin, 1988).

E. coli TP2504 : F', *xyl argH1 ilvA zcf-229::Tn10 ptsG22* (De Reuse et Danchin, 1988) (Don de H. De Reuse, Paris).

Ces souches ont été utilisées pour tester l'effet de l'absence du transporteur EIIBC^{Glc} sur l'expression des gènes lors de notre étude.

E. coli TP2006 : F' *xyl lacΔX74 Δcya glp-8306* (Roy et Danchin, 1982).

E. coli TP2862 : F', *xyl argH1 ilvA aroB Δ(crr) Km^R* (Levy *et al.*, 1990).

E. coli TP2512 : F⁻ *xyl argH1 ilvA zcf-229::Tn10 ptsG22 Δcya glp8306* (De Reuse et Danchin, 1988).

Ces souches ont été utilisées pour tester l'effet du complexe CRP-AMPc, de la protéine EIIA^{Glc} et de la double délétion *ptsG* et *cya* sur la transcription des gènes lors de notre étude (Don de A. Kolb, Paris).

E. coli C600 : *thr-1, leuB6, lacY1, thi-1, tonA21, supE44, e14-(mcrA)* (Raleigh *et al.*, 1988).

E. coli C600 *fis::kan*.

Ces souches ont été utilisées afin d'étudier l'influence de la protéine Fis sur l'efficacité des promoteurs *gapA* P1 et *yeaA* P1 et P2. Elles nous ont été fournies par J. Rouvière-Yaniv (Paris).

E. coli JTT1 : F⁻ *pyrF gal25 nirA strA19 ura*.

E. coli JTT1 *topA10 ura*.

E. coli JTT1 *gyrB203 (ts) gyrB221 (Nov^R) tna::Tn 10 ura*.

Ces souches ont été utilisées pour étudier l'effet du taux de surenroulement sur l'efficacité du promoteur *gapA* P1. Elles nous ont été fournies par J. Rouvière-Yaniv (Paris).

E. coli BL21 DE3 : B, F⁻, *ompT, hsdS_B (r_b⁻ m_b⁻), gal, dcm, λ(DE3)*. Cette souche a été utilisée pour produire la protéine Hfq. Elle possède une copie du gène codant l'ARN pol du bactériophage T7, placé sous contrôle du promoteur *lacUV5* inductible par l'IPTG et intégré dans le chromosome par l'intermédiaire du prophage λ(DE3).

II.) Milieux de culture.

<u>Milieu riche LB (Luria Bertani) :</u>	Tryptone	20 g/l
	Extrait de levure	5 g/l
	NaCl	10 g/l

Le pH est ajusté à 7,5 avec NaOH 10 N. Le milieu de culture est ensuite autoclavé 20 minutes à 121°C sous une pression de 1 bar.

<u>Milieu minimum M63 :</u>	KH ₂ PO ₄	0,1 M
	(NH ₄) ₂ SO ₄	15 mM
	MgSO ₄ , 7H ₂ O	0,8 mM
	FeSO ₄ , 7H ₂ O	2 μM

Le pH est ajusté à 7,0 par KOH 10 N et le milieu de culture est autoclavé 30 minutes à 0,5 bar.

Le milieu M63 est ensuite complétement avec différentes sources de carbone :

<u>Milieu M63 complétement :</u>	Casaminoacides	2,5 %
	Thiamine	0,2 %
	Source de carbone (20 %)	0,2 %

Différentes sources de carbone ont été utilisées : glucose, pyruvate.

Milieux solides : 17 g d'agar par litre de milieu de culture ont été ajoutés avant autoclavage.

Les différents milieux de culture peuvent être supplémentés avec des antibiotiques :

→ Ampicilline 100 mg/l.

→ Kanamycine 100 mg/l.

III.) Plasmides et bactériophages utilisés.

Plasmides:

- pBS SK+ (Clontech). Ce plasmide a été utilisé pour les clonages de gènes étudiés.
- pBS::*EcogapA* (Fig. 75). Ce plasmide contient le gène *gapA* sous contrôle de sa région promotrice (Charpentier, 1994).
- pRLG 770 (Ross *et al.*, 1990). Ce plasmide nous a été fourni par R. Gourse (University of Wisconsin).
- pMosBlue et pT-Adv. Ces plasmides ont été utilisés pour le clonage de fragments synthétisés par la technique PCR (Amersham Pharmacia Biotech et Clontech, respectivement).

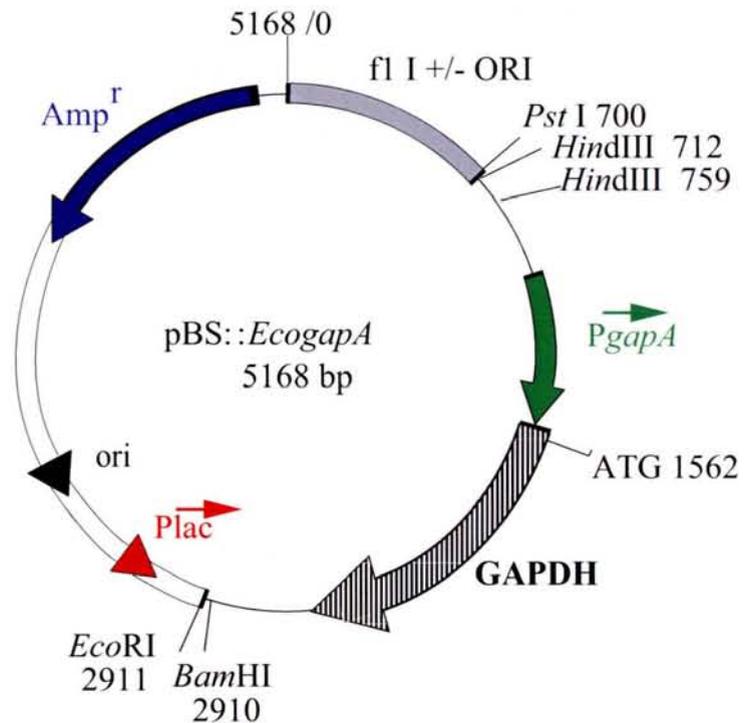


Figure 75 : Carte de restriction du plasmide pBS::EcogapA (Charpentier et Branlant, 1994 ; Charpentier, 1994).

La flèche grisée représente la région codante de la GAPDH et la flèche verte représente la région promoteur du gène *gapA* (*PgapA*).

Amp^R : gène de résistance à l'ampicilline.

ori : origine de la réplication du plasmide colE1.

Plac : promoteur de l'opéron lactose.

fl I +/- ORI : origine de réplication du phage filamenteux fl.

- pRS415 (Simons *et al.*, 1987). Ce plasmide contient une partie de l'opéron lactose, dont l'ORF *lacZ* précédée de sa séquence RBS. Les sites uniques de restriction, permettant le clonage des gènes, sont localisés entre des signaux forts de terminaison de la transcription et la séquence RBS de l'ORF *lacZ*. Il est utilisé pour obtenir des fusions transcriptionnelles entre les promoteurs à étudier et l'ORF *lacZ*.
- pRS414 (Simons *et al.*, 1987). Ce plasmide est similaire au plasmide pRS415. Il ne contient cependant pas la séquence RBS de l'ORF *lacZ*. Les sites uniques de restriction sont localisés au niveau des premiers codons de l'ORF *lacZ*. Il est utilisé pour obtenir des fusions traductionnelles entre les gènes à étudier et l'ORF *lacZ*.
- pUC19 (Biolabs). Ce plasmide a été utilisé pour le clonage du gène *csrA* et des séquences codant les ARNnc CsrB et RyhB.
- pET21b-hfq. Ce plasmide sert à la production de la protéine Hfq. L'ADN codant la protéine Hfq a été cloné en aval d'un promoteur reconnu par l'ARN pol du bactériophage T7 (Don de G. Storz, NIH, USA).

Bactériophage : λ RS45 (Simons *et al.*, 1987) (*imm21*, *ind+*, *bla'*-*lacZSC*). Ce phage permet l'obtention de phages recombinants portant des fusions entre le gène à étudier et l'ORF *lacZ*. Il contient une partie de l'opéron lactose, ainsi qu'une région du gène *bla*. La recombinaison homologue s'effectue au niveau de ces régions et des régions similaires portées par les plasmides pRS415 et pRS414. Il permet d'obtenir une copie unique de ces fusions dans le génome bactérien, puisque ce phage peut entrer en cycle lysogène.

IV.) Tampons et solutions à usage courant.

TE : EDTA 1 mM, Tris-HCl 10 mM pH 8,0.

TBE 1X : acide borique 89 mM, EDTA 2mM, Tris-Borate 89 mM pH 8,3.

Solution de dépôt pour gels d'agarose : EDTA 125 mM, SDS 2 %, glycérol 25 %, bleu de bromophénol 0,12 %.

Solution de dépôt pour gel de polyacrylamide dénaturant (bleu de formamide) : EDTA 20 mM, bleu de bromophénol 0,05 %, bleu de xylène cyanol 0,05% dans de la formamide désionisée.

Bleu de protéine : Tris HCl 80 mM pH 6.8, SDS 3 %, Bleu de bromophénol 10 mg/ml, DTT 100 mM, glycérol 10 %.

Partie 2 : Méthodes.

I.) Extraction et analyse de l'ADN.**I.1) Extraction d'ADN plasmidique.**

La préparation d'ADN plasmidique à petite échelle a été réalisée par la technique de lyse alcaline (Sambrook *et al.*, 1989) sur des cultures bactériennes de 2 ml arrêtées en phase stationnaire de croissance. Pour la production d'ADN plasmidique en grande quantité, le kit Quick Flow Midi Kit Protocol (Qiagen) a été utilisé.

La qualité des préparations est vérifiée par fractionnement des produits dans un gel d'agarose à 0,8% (p/v). Le gel est préparé dans du tampon TBE 0,5x contenant du bromure d'éthidium à une concentration finale de 0,5 µg/ml. Le tampon TBE 0,5x est également utilisé comme tampon d'électrophorèse. On ajoute aux échantillons 1/5ème à 1/10ème de volume de bleu de dépôt agarose avant de les déposer sur gel, en parallèle avec un marqueur de taille. La migration est effectuée à voltage constant (5 V/cm).

I.2) Marquage des oligonucléotides.

Les oligonucléotides achetés purifiés à la société Eurogentec sont marqués à l'extrémité 5' par du phosphore 32 radioactif. La réaction est catalysée par la T4 polynucléotide kinase en présence de [γ -³²P]ATP.

Le mélange réactionnel est le suivant :

[γ - ³² P]ATP (10 mCi/ml, 3000 Ci/mmol) (Amersham)	50 µCi
Oligonucléotide	100 ng
Tampon Enzyme 10X (MgCl ₂ 0,1 M; DTT 0,05 M; Tris-HCl 0,7 M pH 7,6)	1 µl
PNK (10 U/µl) (Biolabs)	1 µl
H ₂ O	qsp 10 µl

Après une incubation de 1 heure à 37°C, le mélange est placé 10 min à 70 °C pour dénaturer l'enzyme.

I.3) Amplification d'ADN par la technique PCR.

Les cycles de synthèse d'ADN à partir d'oligonucléotides amorces et d'ADN polymérase de *Thermus brockianus* (DyNAzyme, Finnzymes) sont réalisés dans les conditions décrites dans le tableau 5.

I.4) Clonage de fragments d'ADN.

I.4.1) Ligation d'un fragment d'ADN issu d'une hydrolyse par enzymes de restriction.

Le vecteur de clonage est hydrolysé par les enzymes de restriction adéquates, puis traité à la phosphatase alcaline (Boehringer) (1 U/ μ g d'ADN) dans le tampon 10 x suivant : Tris-HCl 500 mM pH 8,5, EDTA 1 mM pendant 1 heure à 37°C. Ce traitement permet d'éviter la religation du vecteur sur lui-même lors de la réaction de ligation. L'enzyme est enfin dénaturée par une incubation de 10 min à 65 °C.

La ligation est catalysée par l'ADN ligase du bactériophage T4. Nous avons travaillé en présence d'un excès de fragment avec un rapport molaire, fragment / vecteur, variant de 4 à 8.

Vecteur hydrolysé et déphosphorylé	50 ng
Fragment d'ADN à insérer, excès molaire par rapport au vecteur	4 à 8 x
Tampon de ligation 10X (Tris-HCl 660 mM pH 7,6, MgCl ₂ 50 mM, DTT 10 mM)	2 μ l
ATP (10 mM)	2 μ l
ADN ligase (1 U/ μ l) (Roche)	1 μ l
H ₂ O	qsp 20 μ l

Le mélange, incubé la nuit à 16°C ou 4 heures à température ambiante, est utilisé directement pour transformer les cellules d'*E. coli* TG1 préalablement rendues compétentes par un traitement au CaCl₂ (50 mM) (Cohen *et al.*, 1972). Après un choc thermique de 30 sec à 42°C, les cellules renfermant un plasmide sont sélectionnées sur un milieu LB + ampicilline.

A- Amplification d'ADN par la méthode PCR

Matrice ADN	10 ng
Oligonucléotide (100 ng)	2 µl
dNTP (20 mM chaque)	0,5 µl
Tampon 10x (MgCl ₂ 15 mM, KCl 500 mM, Tris-HCl 100 mM, pH 8,3)	5 µl
DyNAzyme (Finnzymes) (1 U/µl)	0,75 µl
H ₂ O	qsp 50µl

L'amplification est réalisée dans un appareil Perkin-Elmer. Les cycles suivant sont réalisés :

cycle	fonction	température	temps	Nombre de cycles
1	dénaturation	94°C	2 minutes	1
2-32	dénaturation	94°C	30 secondes	30
	Hybridation	48°C	30 secondes	
	élongation	72°C	45 secondes	
33	élongation	72°C	10 minutes	1
34	stockage	4°C		

B- Amplification d'ADN à partir d'une colonie par la méthode PCR

Mélange réactionnel :

Oligonucléotide (20 µM)	1,25 µl
Tampon 10x	2,5 µl
dNTP	0,25 µl
DyNAzyme (Finnzymes) (1 U/µl)	0,25 µl
H ₂ O	qsp 25 µl

La colonie est ensuite ajoutée dans ce mélange réactionnel. Les cycles suivants sont ensuite réalisés :

Cycle	fonction	température	temps	Nombre de cycles
1	dénaturation	94°C	40 secondes	1
2-32	dénaturation	94°C	30 secondes	30
	Hybridation	42°C	30 secondes	
	élongation	72°C	30 secondes	
33	élongation	72°C	10 minutes	1
34	stockage	4°C		

Tableau 5 : Utilisation de la technique PCR pour l'amplification d'ADN

I.4.2) Ligation d'un fragment d'ADN synthétisé par la technique PCR.

Les fragments d'ADN db synthétisés par l'ADN pol de *T. brockianus* possèdent une adénosine supplémentaire aux extrémités 3'. Le clonage de ces fragments s'effectue dans un plasmide linéarisé appelé plasmide "T", possédant une thymine excédentaire sur chaque extrémité 5'. Les kits de clonage pMosBlue (Amersham Pharmacia Biotech) et pT-Adv (Clontech) basés sur cette technique ont été utilisés.

I.4.3) Vérification de l'insertion.

Pour vérifier l'insertion du fragment, deux techniques peuvent être utilisées : l'analyse par restriction et séquençage du vecteur après extraction des plasmides par la technique de lyse alcaline, ou la technique PCR appliquée directement sur colonie. Cette dernière méthode permet de tester rapidement si le fragment d'ADN a été inséré dans le vecteur. Un transformant est mis en suspension dans le mélange réactionnel dont la composition est donnée dans le tableau 5. Différents couples d'oligonucléotides peuvent être utilisés pour vérifier si le fragment est inséré. Après fractionnement en gel d'agarose des produits d'amplification, la taille des produits d'amplification permet de sélectionner le vecteur recombinant recherché.

I.5) Séquençage d'ADN.

Le séquençage de l'ADN double brin, dissocié en simple brin par une étape préalable de dénaturation à 94°C, est réalisé selon la méthode de Tabor et Richardson (1987). Le principe repose sur l'utilisation d'un oligonucléotide de synthèse qui, hybridé en 3' du fragment à séquencer, sert d'amorce à la synthèse d'un ADN complémentaire par l'ADN polymérase du phage T7. La synthèse obtenue par allongement de l'amorce est effectuée en présence de désoxynucléotides (dNTP) et de didésoxy nucléotides (ddNTP). L'incorporation de ddNTP arrête la synthèse, l'arrêt est statistique en fonction de la concentration relative des dNTP/ddNTP (Tabor et Richardson, 1987).

I.6) Technique de recombinaison de fragment d'ADN à l'aide du bactériophage λ .

Principe : cette technique permet d'insérer dans le chromosome d'*E. coli*, des constructions du type fusions transcriptionnelles ou traductionnelles, à l'état de copie unique. Le site d'insertion est le site *att* du phage λ . Ces constructions sont réalisées dans les plasmides pRS414 ou pRS415 en amont de l'ORF *lacZ*. Des phages λ recombinants qui portent les fusions sont ensuite produits et sont utilisés pour infecter des cellules d'*E. coli*. Les cellules qui portent la fusion insérée à l'état de copie unique sur le chromosome sont ensuite sélectionnées.

I.6.1) Obtention et production de bactériophages recombinants.

La souche *E. coli* IBPC5321, *recA*⁺, est transformée par les plasmides pRS414 ou pRS415 contenant les fusions. Ces cellules sont ensuite infectées par le phage λ RS45. La réaction suivante est réalisée :

100 μ l de préculture en phase stationnaire des cellules transformées par les plasmides pRS414 ou pRS415 et 10^7 phages λ RS45 sont placés 10 min à 30 °C sans agitation. Cette incubation permet l'adsorption des phages sur les cellules.

Puis, 25 ml de milieu de culture LBMM (LB additionné de maltose à 0,2 % et de MgSO₄ 10 mM final) sontensemencés par le mélange précédent et sont incubés à 37°C sous agitation jusqu'à l'obtention d'une lyse cellulaire. La lyse est complétée par l'addition de 8 gouttes de chloroforme, suivie d'une incubation de 15 min à 37°C sous agitation. Les cultures sont centrifugées à 4000 g durant 15 min à 4°C. Les surnageants récupérés sont dilués 10^3 , 10^4 et 10^6 fois dans du tampon MgSO₄ 10 mM.

L'isolement et la titration des phages recombinants se fait de la façon suivante :

100 μ l de ces dilutions sont mélangés à 100 μ l d'un stock de cellules de la souche TG1 (ce stock cellulaire est obtenu par concentration d'un facteur 2,5 dans du tampon MgCl₂ 10 mM, d'une culture de cette souche en phase stationnaire). Le mélange est incubé 10 min à 30°C, sans agitation, placé dans 3 ml de gélose molle puis étalé sur boîtes de Pétri LBMM contenant le substrat X-gal (40 μ g/ml). Les phages recombinants sont identifiés par leur capacité à

produire de la β -galactosidase. Ils sont purifiés par des étalements successifs sur boîtes LBMM/X-gal, sur lesquelles 100 μ l du stock de cellules TG1 ont été étalés.

L'amplification et la production des phages recombinants est alors réalisée en infectant la souche d'*E. coli* TG1 non transformée. Les conditions opératoires sont similaires à celles utilisées pour l'obtention des phages recombinants. Le titre des surnageants est déterminé par la méthode préalablement décrite dans cette partie.

I.6.2) Obtention de souches lysogènes.

Les cellules TG1 ou IBPC5361 sont infectées par les phages recombinants de la façon suivante :

5 μ l de stock cellulaire
2,5x10⁷ phages recombinants
qsp 1ml LBMM

Après une nuit d'incubation à 37°C, sans agitation, les cellules infectées sont diluées 10⁴ fois dans du tampon MgSO₄ 10 mM et 100 μ l sont étalés sur boîtes LB/X-gal. Les cellules lysogènes sont identifiées par leur capacité à produire de la β -galactosidase non accompagnée de lyse cellulaire. Elles sont purifiées par des étalements successifs sur boîtes LB/X-gal.

I.6.3) Sélection des cellules monolysogènes.

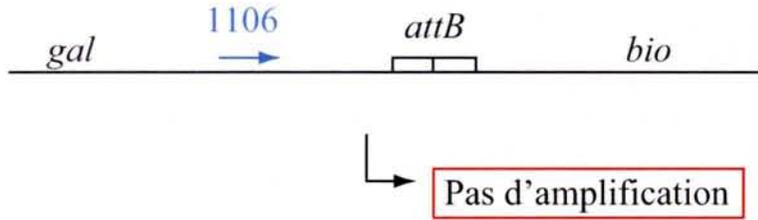
Pour vérifier la monolysogénie des souches, un test PCR est réalisé (Powell *et al.*, 1994) de la façon suivante :

Une colonie de chaque souche lysogène est traitée dans les conditions décrites dans le tableau 5B avec trois oligonucléotides comme indiqué dans la figure 76.

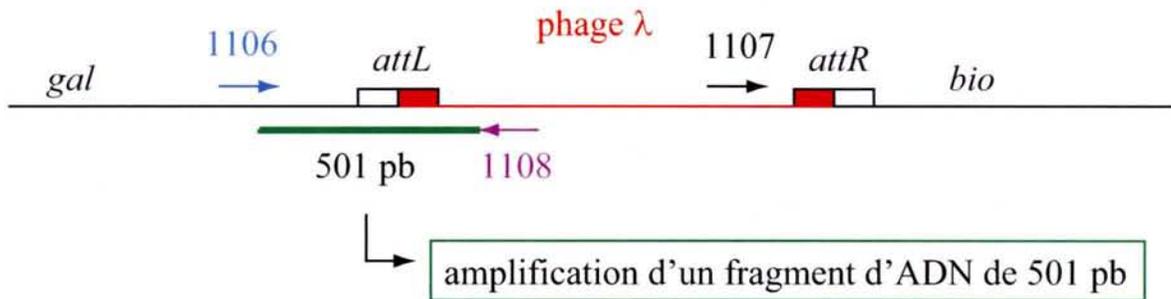
Les profils d'amplification suivants peuvent être obtenus suivant les cellules :

1 fragment de 501 pb	les cellules sont monolysogènes.
2 fragments de 501 pb et de 379 pb	les cellules sont multilyso-gènes.
aucun fragment	les cellules ne sont pas lysogènes.

cellule non lysogène



cellule monolysogène



cellule multilysoène

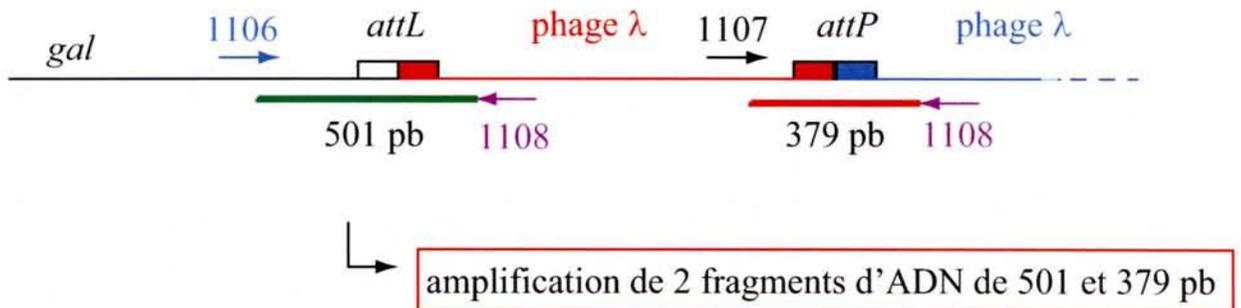


Figure 76 : Principe du test d'amplification par PCR permettant d'identifier les cellules monolysogènes (d'après Powell *et al.*, 1994).

Les sites *att* du chromosome bactérien sont indiqués en blanc, et ceux des phages en couleur. Les oligonucléotides permettant l'amplification des régions *att* en produisant des fragments de tailles différentes sont indiqués par des flèches de couleur.

oligonucléotide 1106 : 5' GAGGTACCAGCGCGGTTTGATC 3'

oligonucléotide 1107 : 5' TTTAATATATTGATATTTATATCATTTCACGTTTCTCGTTC 3'

oligonucléotide 1108 : 5' ACTCGTCGCGAACCGCTTTC 3'

I.7) Mutagenèse dirigée par PCR.

Les variants ont été produits par la technique de mutagenèse par PCR en utilisant le kit QuikChange site-directed mutagenesis (Stratagene). Cette technique repose sur le fait que l'ADN plasmidique qui sert de matrice à la PCR a été produit dans une souche d'*E. coli* XL1 blue qui possède un système de méthylation actif. Ce plasmide qui porte par conséquent des méthylations, est amplifié par l'ADN pol de *Pyrococcus furiosus* (*Pfu*) à partir de deux oligonucléotides complémentaires l'un de l'autre et contenant tous les deux la mutation à introduire. Après amplification, la matrice est dégradée par l'enzyme de restriction *Dpn* I qui reconnaît et clive les séquences des sites méthylés (5' GAMTC 3'). Ces sites ne sont présents que sur le plasmide parental produit chez *E. coli* mais pas sur l'ADN amplifié par PCR. Les produits PCR sont ensuite utilisés pour transformer des bactéries qui produiront des plasmides qui portent la mutation.

Le mélange réactionnel est le suivant :

Plasmide matrice	10 ng
Tampon de PCR 10X (Tris-HCl 200 mM pH 8,8, MgSO ₄ 20 mM, KCl 100 mM, (NH ₄) ₂ SO ₄ 100 mM, Triton X100 1 %, BSA 1 mg/ml)	5 µl
Amorce 5'	100 ng
Amorce 3'	100 ng
Mélange des 4 dNTP (1,25 mM chacun)	8 µl
<i>Pfu</i> (2,5 U/µl) (Stratagène)	1 µl
H ₂ O	qsp 50 µl

Après un cycle de dénaturation de 5 minutes à 95°C, l'amplification se déroule en 18 cycles de 30 secondes à 95°C, 1 minute à 55°C et 10 minutes à 68°C. Les produits PCR sont visualisés sur gel d'agarose à 0,8 %. La matrice est ensuite digérée par 10 U d'enzyme *Dpn* I pendant 1 heure à 37°C. Enfin, les bactéries *E. coli* XL1 blue, rendues compétentes par traitement au CaCl₂ (50 mM) (Sambrook *et al.* 1989), sont transformées avec 1 ou 4 µl de produit PCR traité par *Dpn* I.

II.) Analyse des ARN.

II.1) Extraction des ARN totaux.

Une préculture en phase stationnaire de croissance est utilisée pour ensemercer 25 ml de milieu de culture de façon à obtenir une DO à 600 nm (DO_{600}) de 0,1. Lorsque la DO_{600} de la culture atteint 0,5 (en milieu de phase exponentielle), les cellules sont centrifugées puis lavées avec du tampon A (EDTA 5 mM, saccharose 25 %, Tris-HCl 50 mM pH 8,0).

Dans le cas d'une étude cinétique de l'évolution des transcrits au cours d'une culture, des échantillons correspondant à 6×10^8 cellules sont prélevés régulièrement au cours de la culture. L'extraction des ARN totaux est réalisée par la méthode basée sur l'emploi du phénol à chaud (65°C) (Aiba *et al.*, 1981). Les ARN extraits sont précipités à l'éthanol et à l'acétate de sodium 0,3 M, puis repris dans de l'eau stérile. Une quantification est réalisée par mesure de l'absorbance à 260 nm avec un appareil Genequant (Pharmacia).

II.2) Technique d'extension d'amorces marquées au ^{32}P par la transcriptase inverse.

Un oligonucléotide marqué en 5' au ^{32}P est hybridé à l'ARN. L'extrémité 3' de cet oligonucléotide sert d'amorce à la transcriptase inverse qui synthétise, en présence des 4 dNTP, un brin d'ADN complémentaire à l'ARN.

II.2.1) Hybridation ARN/oligonucléotide.

La quantité d'ARN total est variable selon la provenance des ARN. Pour des ARN totaux extraits de cellules non transformées, 20 μg sont incubés en présence de 4 ng d'oligonucléotide marqué (le protocole de marquage est détaillé ci-dessus), alors que pour des ARN totaux extraits de cellules transformées par un plasmide multicopie, 5 μg sont placés en présence de 4 ng d'oligonucléotide marqué. Ces mélanges sont réalisés dans le tampon de la transcriptase inverse d'AMV (fourni avec l'enzyme) et dans un volume final de 15 μl . Après une incubation de 10 min à 65°C , ils sont placés à température ambiante pendant 10 min.

II.2.2) Synthèse du brin complémentaire.

Elle est réalisée en ajoutant : 1 μ l (soit 0,1 U) de transcriptase inverse d'AMV (Life Science), 1,5 μ l d'un mélange de dNTP (1,25 mM chaque), 0,5 μ l de tampon enzyme 10X et 1,5 μ l d'H₂O afin de se placer dans un volume final de 20 μ l. Les réactions sont effectuées à 42°C pendant 30 minutes, puis sont arrêtées par l'addition de 100 μ l d'acétate de sodium 0,3 M et 500 μ l d'éthanol absolu. Après centrifugation, le précipité est lavé avec 200 μ l d'éthanol 70 %, puis séché.

II.2.3) Fractionnement des ADNc obtenus.

Les précipités sont repris dans 8 μ l de bleu de formamide. Après chauffage 2 min à 95°C, les ADNc obtenus sont déposés sur un gel de polyacrylamide dénaturant (polyacrylamide 5%, urée 8M, TBE 1X) ayant subi une pré-électrophorèse d'une heure à 40 mA. La migration est réalisée à 45 V/cm, puis le gel est transféré sur papier Whatman 3MM, recouvert d'un film plastique et séché. Les bandes de radioactivité peuvent être quantifiées par utilisation du logiciel Image Quant (version 5.2) de l'appareil PhosphorImager (Molecular dynamics).

II.3) Protocoles de transcription *in vitro*.

II.3.1) Production d'ARN froids ou uniformément marqués par transcription *in vitro* avec l'ARN pol T7.

Cette transcription peut être réalisée à partir d'un fragment d'ADN obtenu par amplification PCR. Dans ce cas, la région codant l'ARN à transcrire a été amplifiée avec un oligonucléotide amorce générant une séquence promoteur reconnue par l'ARN polymérase du phage T7. Des ribonucléotides rUTP [α -³²P] sont utilisés pour la transcription de façon à produire des ARN uniformément marqués.

Le fragment d'ADN amplifié par PCR est ensuite soumis à une extraction par un mélange phénol-chloroforme. L'ADN est ensuite précipité à l'éthanol en présence d'acétate d'ammonium (2 M) puis lavé à l'éthanol 70 %. L'ADN est ensuite repris dans de l'eau stérile MilliQ et sa concentration est ajustée à 0,5 μ g/ μ l.

Protocole :

ADN (0,5 µg/µl)	1 µl
Tampon de transcription 3x (Tris-HCl 120 mM pH 8,0, MgCl ₂ 60 mM, Spermidine 3 mM, DTT 15 mM, triton 0,03%, PEG 8000 240 mg/ml).	10 µl
rNTP (ATP, CTP, GTP 20 mM)	6 µl
rUTP (5 mM)	2,4 µl
[α- ³² P] UTP (800 Ci/mmol; 10 mCi/ml) (ICN)	2,4 µl
RNasin (31 U/µl) (Amersham Biosciences)	0,5 µl
ARN pol T7 (purifiée au Laboratoire)	1 µl
BSA (2 mg/ml)	3 µl
H ₂ O	qsp 30 µl

La réaction est incubée à 37°C pendant un temps minimum de 2 heures puis 10 U de DNase I sont ajoutés. Le mélange est à nouveau incubé à 37°C pendant 30 min.

Les transcrits sont purifiés sur gel de polyacrylamide dénaturant (polyacrylamide 12%, urée 8M, TBE 1X), puis élués une nuit à 4°C dans du tampon d'éluion à forte concentration saline (Tris-HCl 10 mM pH 7,5, NaCl 300 mM, EDTA 1 mM, SDS 1 %). Les transcrits sont ensuite précipités à l'éthanol puis lavés à l'éthanol 70 %. Les transcrits sont ensuite repris dans un de l'eau stérile. La quantité de transcrits radioactifs obtenue est déterminée par mesure de la radioactivité contenue dans 1 µl d'ARN au compteur à scintillation (Beckman).

II.3.2) Analyse de l'efficacité d'un promoteur par transcription *in vitro*.

Les plasmides dérivés du pRLG770 contenant les différents promoteurs insérés dans la cassette de clonage servent de matrice pour la synthèse d'ARN par l'ARN pol Eσ⁷⁰ d'*E. coli*. Les ARN sont uniformément marqués grâce à l'incorporation d'UTP radioactif à raison de 5 µCi par réaction. Le mélange (ATP : 200 µM, GTP : 5 µM à 1,2 mM, CTP : 200 µM, UTP non marqué : 10 µM) contenant les autres nucléotides a été utilisé.

Protocole :

Plasmide (20 nM)	1 µl
Tampon TIV1 10x	1 µl

(MgCl ₂ 10 mM, DTT 1 mM, BSA 0,1 mg/ml, KCl 170 mM, Tris-acétate 40mM pH 7,9)	
ARN polymérase Eσ ⁷⁰ (Roche) (1 U/μl)	0,25 μl
Mélange de nucléotides	1 μl
UTP [α - ³² P] (800 Ci/mmoles, 10 μCi/μl) (Amersham)	0,5 μl
H ₂ O stérile	qsp 10 μl

La réaction est déclenchée par l'addition de l'ARN pol Eσ⁷⁰. Le mélange est incubé 15 min à 22°C. La réaction est arrêtée par ajout de 10 μl de solution d'arrêt (Urée 8 M, EDTA 10 mM, SDS 1 %, TBE 2x, bleu de bromophénol 0,05 %, bleu de xylène cyanol 0,025 %).

La radioactivité contenue dans chaque échantillon après transcription est mesurée (en Cerenkov) au compteur à scintillation (Beckman). Pour chaque échantillon, une fraction aliquote correspondant à la même quantité de radioactivité est ensuite déposée sur un gel de polyacrylamide dénaturant à 5 % (acrylamide/bisacrylamide 25 % (p/v 23,75/1,25), urée 8 M). Le dépôt est effectué après que les échantillons aient été dénaturés 40 secondes à 96°C, puis plongés dans la glace. La migration est effectuée en tampon TBE 1x, à 45 V/cm pendant 40 minutes et à une température de 48°C. Le gel est ensuite séché et placé en exposition dans une cassette de PhosphorImager (Molecular Dynamics) pendant 2 heures. La radioactivité contenue dans les différentes bandes de gel est ensuite estimée par l'utilisation du logiciel Image Quant (Molecular Dynamics).

II.4) Etude des complexes [ARN-Hfq].

II.4.1 Formation des complexes [ARN-Hfq].

Les complexes [ARN-Hfq] sont formés dans du tampon Hfq 1x (Tris-HCl 10 mM pH 8; MgCl₂ 10 mM; NaCl 50 mM; KCl 50 mM) en incubant 30 min à 37°C, 100 fmol d'ARN marqué uniformément (5000 cpm) en présence de concentrations croissantes en protéine Hfq, de 20 ng d'ARNt de levure (Roche Molecular Biochemicals) et 0,5 μl de RNasin (Roche Molecular Biochemicals, 29,1 U/μl).

II.4.2 Fractionnement des complexes [ARN-Hfq].

L'électrophorèse en gel natif

Principe: La formation du complexe [ARN-protéine] est visualisée par un retard de migration par rapport à l'ARN libre.

Méthode: 1 µl de bleu Hfq (TBE 1x; Glycérol 50% (p/v); bleu de bromophénol 0,1% (p/v); xylène cyanol 0,1% (p/v)) sont ajoutés au mélange réactionnel ayant servi à former le complexe. L'échantillon est ensuite déposé sur un gel de polyacrylamide natif, acrylamide-bisacrylamide (38:2) 6%, en présence de TBE 0,5x. L'électrophorèse se fait à 15 V/cm et à température ambiante. Le gel est ensuite séché puis exposé au PhosphorImager (Molecular Dynamics).

II.5) Analyse par coupure à la RNase T1.

Les hydrolyses ont été réalisées dans des conditions ménagées en présence d'ARNt pour que statistiquement, chaque molécule d'ARN radiomarquée soit coupée une seule fois. La position de ces sites de coupures est identifiée par électrophorèse en gel de polyacrylamide dénaturant.

La première étape est de déphosphoryler l'extrémité 5' de l'ARN par la phosphatase alcaline puis d'ajouter un phosphate radioactif par la T4 polynucléotide kinase (le protocole est le même que celui utilisé pour phosphoryler l'extrémité 5' d'un oligonucléotide). Les transcrits marqués ont été purifiés sur gel de polyacrylamide dénaturant 10 % (urée 8M, TBE 1X), élués dans un tampon contenant des sels et précipités avec de l'éthanol puis lavés par de l'éthanol 70 %. Les ARN sont repris dans un volume d'eau stérile et la radioactivité est mesurée par un compteur à scintillation (Beckman).

Coupure à la RNase T1 :

ARN (20000 cpm)	1 µl
ARNt (2 µg/µl)	1 µl
Tampon (10 x) (Tris-HCl 50 mM, MgCl ₂ 10 mM, KCl 150 mM)	2 µl
H ₂ O	qsp 20 µl

Le mélange est incubé 20 min à température ambiante et différentes quantités de RNase T1 sont ajoutées (de 1 à 3 unités). Le mélange est incubé 10 min à 20°C et la réaction est stoppée par ajout de 20 µg d'ARNt, suivi immédiatement d'une extraction par un mélange phénol/chloroforme. Après précipitation à l'éthanol en présence d'AcNa 0,3 M, les ARN sont repris dans du bleu de formamide et déposés sur gel de polyacrylamide dénaturant (polyacrylamide 7%, urée 8M, TBE 1X).

III) Analyses des protéines cellulaires.

III.1) Cultures et sonication des bactéries.

Les cultures sont arrêtées par centrifugation et les culots cellulaires sont repris dans 1/10ème de volume de tampon de sonication froid (EDTA 2 mM; Tris-HCl 20 mM pH 7,0). La sonication est alors réalisée à 0°C sous une puissance de 20 W durant 2 à 3 min par période d'1 seconde d'ultrason et 1 seconde d'arrêt. Les débris cellulaires sont éliminés par centrifugation et les protéines sont quantifiées par mesure de l'absorbance du surnageant à 280 nm. Les protéines peuvent alors être visualisées par fractionnement dans un gel de polyacrylamide dénaturant (Sambrook *et al.*, 1989).

III.2) Mesure des activités enzymatiques.

III.2.1) Mesure de l'activité β -galactosidase.

Les cellules sont lysées par traitement au toluène. La β -galactosidase hydrolyse le O-Nitrophényl β -D-galactopyranoside (ONPG) en libérant un composé chromophore, absorbant à 420 nm qui permet ainsi la détermination du taux de β -galactosidase produite par les cellules (Test de Miller) (Miller, 1972).

Test enzymatique :

Les échantillons de cultures sont prélevés lorsque les cultures atteignent 0,2 et 0,4 U DO_{650nm}/ml . Un volume variable de culture est prélevé et complété à 1 ml avec du tampon Z (Na_2HPO_4 60 mM, NaH_2PO_4 40 mM, KCl 10 mM, $MgSO_4$ 1 mM, β -mercaptoéthanol 50 mM). Une goutte de toluène est ajoutée et les échantillons sont agités vigoureusement pour

lyser les cellules. Après évaporation du toluène à 37°C, 200 µl d'ONPG (4 mg/ml dans du tampon Na₂HPO₄ 0,1 M) sont ajoutés. Le mélange est incubé à 28°C et la réaction est arrêtée, lors de l'obtention d'une couleur jaune intense, par ajout de 0,5 ml de Na₂CO₃ 1 M. Les DO à 420 nm et à 550 nm sont mesurées. Les DO₅₅₀ permettent de corriger les valeurs de DO₄₂₀, faussées par la présence de débris cellulaires.

L'activité β-galactosidase, exprimée en activité Miller, est calculée par la formule suivante (Miller, 1972) :

$$AS=1000x[DO_{420}-(1,75xDO_{550})] / [DO_{650}xTempsxVol]$$

Temps : temps de réaction à 28°C (min)

Vol : volume de la prise d'essai (ml).

III.2.2) Mesures d'activité GAPDH.

La GAPDH catalyse la réaction de conversion du glycéraldéhyde-3-phosphate (G3P) en 1,3-diphosphoglycérate (1,3-DPG). Lors de la réaction, le NAD⁺ est réduit en NADH + H⁺. L'activité enzymatique est donc suivie par apparition du NADH en fonction du temps. Cette apparition est mesurée par la variation de la DO à 340 nm (test de Ferdinand, 1964).

La réaction est réalisée dans le tampon Ferdinand (K₂HPO₄ 50 mM, EDTA 0,2 mM, Triéthanolamine 40 mM, pH ajusté à 9,2 par H₃PO₄²⁻) en présence de NAD⁺ et du substrat G3P. La réaction est amorcée par ajout de l'extrait protéique.

III.2.2.1) Préparation du substrat et du cofacteur.

→ Préparation du D-G3P : 200 mg de diéthylacétal D-G3P (DL-glycéraldéhyde-3-phosphate diéthylacétal, sel de monobarium ; Sigma Chemical) et 3 g de résine DOWEX 50 (activée) dans 5 ml d'eau sont portés à 100°C au bain marie pendant 3 minutes. Après décantation de la résine, le surnageant est prélevé puis filtré. La solution est dosée précisément par un test enzymatique avec de la GAPDH purifiée en présence de NAD⁺ et d'arsenate (rendant la réaction irréversible).

→Préparation du NAD⁺ : le NAD⁺ (free acid, grade I ; Boehringer) est dissous dans un tampon EDTA 20 mM, Tris-HCl 500 mM, pH 8,0. La concentration initiale est déterminée par mesure de l'absorbance à 259 nm et est ensuite ajustée à 100 mM.

III.2.2.2) Test enzymatique et calcul de l'activité enzymatique.

Le milieu réactionnel suivant est préparé :

G3P	1 mM
NAD ⁺ (100 mM)	5 µl
Tampon Ferdinand	qsp 500 µl

L'activité spécifique de la GAPDH est mesurée en µmoles de NADH / minute / DO₂₈₀.

L'activité spécifique (AS) est calculée à partir de :

$$AS = \frac{\text{Pente } (\Delta DO_{340}/\text{min}) \times \text{dilution} \times 500}{\text{Volume de l'échantillon } (\mu\text{l}) \times 6,22 \times DO_{280}}$$

Le facteur de 6,22 est le coefficient d'extinction molaire du NADH à 340 nm ($6220 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

III.3) Analyse des protéines par gel SDS-PAGE.

Les protéines ont été fractionnées en conditions dénaturantes selon la méthode décrite par Laemmli (1970) sur un appareillage Mini Protean (Biorad). Les protéines ont été colorées en utilisant un tampon contenant du bleu de Commassie R-250.

Bibliographie

S.C.D. - U.H.P. NANCY 1
BIBLIOTHÈQUE DES SCIENCES
Rue du Jardin Botanique
54600 VILLERS-LES-NANCY

A

- Abramson J, Smirnova I, Kasho V, Verner G, Iwata S, Kaback HR. 2003. The lactose permease of *Escherichia coli*: overall structure, the sugar-binding site and the alternating access model for transport. FEBS Lett., 555: 96-101.
- Ades SE. 2004. Control of the alternative sigma factor σ^E in *Escherichia coli*. Curr. Opin. Microbiol., 7:157-62.
- Adler J, Templeton B. 1967. The effect of environmental conditions on the motility of *Escherichia coli*. J. Gen. Microbiol., 46:175-84.
- Aiba H. 1983. Autoregulation of the *Escherichia coli* *crp* gene: CRP is a transcriptional repressor for its own gene. Cell, 32:141-9.
- Aiba H. 1985. Transcription of the *Escherichia coli* adenylate cyclase gene is negatively regulated by cAMP-cAMP receptor protein. J. Biol. Chem., 260:3063-70.
- Aiba H, Adhya S, de Crombrughe B. 1981. Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. J. Biol. Chem., 256:11905-10.
- Aiba H, Matsuyama S, Mizuno T, Mizushima S. 1987. Function of *micF* as an antisense RNA in osmoregulatory expression of the *ompF* gene in *Escherichia coli*. J. Bacteriol., 169: 3007-12.
- Aiba H, Mizuno T, Mizushima S. 1989. Transfer of phosphoryl group between two regulatory proteins involved in osmoregulatory expression of the *ompF* and *ompC* genes in *Escherichia coli*. J. Biol. Chem., 264:8563-7.
- Aiba H, Nakasai F, Mizushima S, Mizuno T. 1989. Evidence for the physiological importance of the phosphotransfer between the two regulatory components, EnvZ and OmpR, in osmoregulation in *Escherichia coli*. J. Biol. Chem., 264: 14090-4.
- Aiba H, Nakasai F, Mizushima S, Mizuno T. 1989. Phosphorylation of a bacterial activator protein, OmpR, by a protein kinase, EnvZ, results in stimulation of its DNA-binding ability. J. Biochem. (Tokyo), 106:5-7.
- Aiyar SE, Gourse RL, Ross W. 1998. Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase α subunit. Proc. Natl. Acad. Sci. USA, 95:14652-7.
- Aiyar SE, McLeod SM, Ross W, Hirvonen CA, Thomas MS, Johnson RC, Gourse RL. 2002. Architecture of Fis-activated transcription complexes at the *Escherichia coli* *rrnB* P1 and *rrnE* P1 promoters. J. Mol. Biol., 316:501-16.
- Aldridge P, Hughes KT. 2002. Regulation of flagellar assembly. Curr. Opin. Microbiol., 5:160-5.
- Alefunder PR, Baldwin SA, Perham RN, Short NJ. 1989. Cloning, sequence analysis and over-expression of the gene for the class II fructose 1,6-bisphosphate aldolase of *Escherichia coli*. Biochem. J., 257:529-34.
- Allison LA, Moyle M, Shales M, Ingles CJ. 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell, 42:599-610.
- Amouyal M, Buc H. 1987. Topological unwinding of strong and weak promoters by RNA polymerase. A comparison between the *lac* wild-type and the UV5 sites of *Escherichia coli*. J. Mol. Biol., 195:795-808.
- Andrews SC, Robinson AK, Rodriguez-Quinones F. 2003. Bacterial iron homeostasis. FEMS Microbiol. Rev., 27:215-37.
- Antoine M, Boschi-Muller S, Branlant G. 2003. Kinetic characterization of the chemical steps involved in the catalytic mechanism of methionine sulfoxide reductase A from *Neisseria meningitidis*. J. Biol. Chem., 278:45352-7.
- Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr. Biol., 11:941-50.
- Arnosti DN, Chamberlin MJ. 1989. Secondary sigma factor controls transcription of flagellar and chemotaxis genes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 86:830-4.
- Arthur TM, Anthony LC, Burgess RR. 2000. Mutational analysis of β' 260-309, a σ^{70} binding site located on *Escherichia coli* core RNA polymerase. J. Biol. Chem., 275:23113-9.
- Arthur TM, Burgess RR. 1998. Localization of a σ^{70} binding site on the N terminus of the *Escherichia coli* RNA polymerase β' subunit. J. Biol. Chem., 273:31381-7.
- Artsimovitch I, Patlan V, Sekine S, Vassilyeva MN, Hosaka T, Ochi K, Yokoyama S, Vassilyev DG. 2004. Structural basis for transcription regulation by alarmone ppGpp. Cell, 117:299-310.
- Auble DT, Allen TL, deHaseth PL. 1986. Promoter recognition by *Escherichia coli* RNA polymerase. Effects of substitutions in the spacer DNA separating the -10 and -35 regions. J. Biol. Chem., 261:11202-6.
- Auble DT, deHaseth PL. 1988. Promoter recognition by *Escherichia coli* RNA polymerase. Influence of DNA structure in the spacer separating the -10 and -35 regions. J. Mol. Biol., 202:471-82.
- Auner H, Buckle M, Deufel A, Kutateladze T, Lazarus L, Mavathur R, Muskhelishvili G, Pemberton I, Schneider R, Travers A. 2003. Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. J. Mol. Biol., 331:331-44.
- Aviv M, Giladi H, Schreiber G, Oppenheim AB, Glaser G. 1994. Expression of the genes coding for the *Escherichia coli* integration host factor are controlled by growth phase, *rpoS*, ppGpp and by autoregulation. Mol. Microbiol., 14:1021-31.
- Ayers DG, Auble DT, deHaseth PL. 1989. Promoter recognition by *Escherichia coli* RNA

polymerase. Role of the spacer DNA in functional complex formation. J. Mol. Biol., 207:749-56.

Azam TA, Ishihama A. 1999. Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. J. Biol. Chem., 274:33105-13.

B

Baichoo N, Heyduk T. 1997. Mapping conformational changes in a protein: application of a protein footprinting technique to cAMP-induced conformational changes in cAMP receptor protein. Biochemistry, 36:10830-6.

Baker CS, Morozov I, Suzuki K, Romeo T, Babitzke P. 2002. CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. Mol. Microbiol., 44:1599-610.

Balandina A, Claret L, Hengge-Aronis R, Rouviere-Yaniv J. 2001. The *Escherichia coli* histone-like protein HU regulates *rpoS* translation. Mol. Microbiol., 39:1069-79.

Baldwin NE, Dombroski AJ. 2001. Isolation and characterization of mutations in region 1.2 of *Escherichia coli* σ^{70} . Mol. Microbiol., 42:427-37.

Ball CA, Osuna R, Ferguson KC, Johnson RC. 1992. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. J. Bacteriol., 174:8043-56.

Barber AM, Zhurkin VB, Adhya S. 1993. CRP-binding sites: evidence for two structural classes with 6-bp and 8-bp spacers. Gene, 130:1-8.

Barbic A, Zimmer DP, Crothers DM. 2003. Structural origins of adenine-tract bending. Proc. Natl. Acad. Sci. USA, 100:2369-73.

Barker MM, Gaal T, Gourse RL. 2001b. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. J. Mol. Biol., 305:689-702.

Barker MM, Gaal T, Josaitis CA, Gourse RL. 2001a. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation *in vivo* and *in vitro*. J. Mol. Biol., 305:673-88.

Bar-Nahum G, Nudler E. 2001. Isolation and characterization of σ^{70} -retaining transcription elongation complexes from *Escherichia coli*. Cell, 106:443-51.

Barnard A, Wolfe A, Busby S. 2004. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. Curr. Opin. Microbiol., 7:102-8.

Barne KA, Bown JA, Busby SJ, Minchin SD. 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended-10' motif at promoters. EMBO J., 16:4034-40.

Bartlett MS, Gourse RL. 1994. Growth rate-dependent control of the *rrnB* P1 core promoter in *Escherichia coli*. J. Bacteriol., 176:5560-4.

Bartlett MS, Thomm M, Geiduschek EP. 2000. The orientation of DNA in an archaeal transcription initiation complex. Nat. Struct. Biol., 7:782-5.

Bauer BF, Kar EG, Elford RM, Holmes WM. 1988. Sequence determinants for promoter strength in the *leuV* operon of *Escherichia coli*. Gene, 63:123-34.

Beatty CM, Browning DF, Busby SJ, Wolfe AJ. 2003. Cyclic AMP receptor protein-dependent activation of the *Escherichia coli* *acsP2* promoter by a synergistic class III mechanism. J. Bacteriol., 185:5148-57.

Bell A, Gaston K, Williams R, Chapman K, Kolb A, Buc H, Minchin S, Williams J, Busby S. 1990. Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. Nucleic Acids Res., 18:7243-50.

Belyaeva T, Griffiths L, Minchin S, Cole J, Busby S. 1993. The *Escherichia coli* *cysG* promoter belongs to the 'extended -10' class of bacterial promoters. Biochem. J., 296:851-7.

Belyaeva TA, Rhodius VA, Webster CL, Busby SJ. 1998. Transcription activation at promoters carrying tandem DNA sites for the *Escherichia coli* cyclic AMP receptor protein: organisation of the RNA polymerase alpha subunits. J. Mol. Biol., 277:789-804.

Benoff B, Yang H, Lawson CL, Parkinson G, Liu J, Blatter E, Ebricht YW, Berman HM, Ebricht RH. 2002. Structural basis of transcription activation: the CAP- α CTD-DNA complex. Science, 297:1562-6.

Berg HC. 2003. The rotary motor of bacterial flagella. Annu. Rev. Biochem., 72:19-54.

Berg OG, von Hippel PH. 1988. Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. J. Mol. Biol., 200:709-23.

Bergstrom LC, Qin L, Harlocker SL, Egger LA, Inouye M. 1998. Hierarchical and co-operative binding of OmpR to a fusion construct containing the *ompC* and *ompF* upstream regulatory sequences of *Escherichia coli*. Genes Cells, 3:777-88.

Beutel BA, Record MT, Jr. 1990. *E. coli* promoter spacer regions contain nonrandom sequences which correlate to spacer length. Nucleic Acids Res., 18:3597-603.

Blatter EE, Ross W, Tang H, Gourse RL, Ebricht RH. 1994. Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell, 78:889-96.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW,

- Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. Science, 277:1453-74.
- Bohm A, Boos W. 2004. Gene regulation in prokaryotes by subcellular relocation of transcription factors. Curr. Opin. Microbiol., 7:151-6.
- Boos W, Shuman H. 1998. Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. Microbiol. Mol. Biol. Rev., 62:204-29.
- Borkovich KA, Alex LA, Simon MI. 1992. Attenuation of sensory receptor signaling by covalent modification. Proc. Natl. Acad. Sci. USA, 89:6756-60.
- Borowiec JA, Zhang L, Sasse-Dwight S, Gralla JD. 1987. DNA supercoiling promotes formation of a bent repression loop in *lac* DNA. J. Mol. Biol., 196:101-11.
- Borukhov S, Nudler E. 2003. RNA polymerase holoenzyme: structure, function and biological implications. Curr. Opin. Microbiol., 6:93-100.
- Borukhov S, Severinov K. 2002. Role of the RNA polymerase sigma subunit in transcription initiation. Res. Microbiol., 153:557-62.
- Boschi-Muller S, Azza S, Branlant G. 2001. *E. coli* methionine sulfoxide reductase with a truncated N terminus or C terminus, or both, retains the ability to reduce methionine sulfoxide. Protein Sci., 10:2272-9.
- Boschi-Muller S, Azza S, Pollastro D, Corbier C, Branlant G. 1997. Comparative enzymatic properties of *GapB*-encoded erythrose-4-phosphate dehydrogenase of *Escherichia coli* and phosphorylating glyceraldehyde-3-phosphate dehydrogenase. J. Biol. Chem., 272:15106-12.
- Boschi-Muller S, Azza S, Sanglier-Cianferani S, Talfournier F, Van Dorsselaar A, Branlant G. 2000. A sulfenic acid enzyme intermediate is involved in the catalytic mechanism of peptide methionine sulfoxide reductase from *Escherichia coli*. J. Biol. Chem., 275:35908-13.
- Botsford JL, Harman JG. 1992. Cyclic AMP in prokaryotes. Microbiol. Rev., 56:100-22.
- Bourret RB, Stock AM. 2002. Molecular information processing: lessons from bacterial chemotaxis. J. Biol. Chem., 277:9625-8.
- Bown JA, Barne KA, Minchin SD, Busby SJW. 1997. Extended -10 promoters. In: Eckstein F, Lilley DMJ, eds. Nucleic Acids and Molecular Biology. Berlin: Springer. pp
- Bowrin V, Brissette R, Tsung K, Inouye M. 1994. The alpha subunit of RNA polymerase specifically inhibits expression of the porin genes *ompF* and *ompC* *in vivo* and *in vitro* in *Escherichia coli*. FEMS Microbiol. Lett., 115:1-6.
- Brandenburg K, Garidel P, Schromm AB, Andra J, Kramer A, Egmond M, Wiese A. 2004. Investigation into the interaction of the bacterial protease OmpT with outer membrane lipids and biological activity of OmpT:lipopolysaccharide complexes. Eur. Biophys. J., 6:6.
- Branlant G, Branlant C. 1985. Nucleotide sequence of the *Escherichia coli gap* gene. Different evolutionary behavior of the NAD⁺-binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. Eur. J. Biochem., 150:61-6.
- Branlant G, Flesch G, Branlant C. 1983. Molecular cloning of the glyceraldehyde-3-phosphate dehydrogenase genes of *Bacillus stearothermophilus* and *Escherichia coli*, and their expression in *Escherichia coli*. Gene, 25:1-7.
- Braun V. 1997. Surface signaling: novel transcription initiation mechanism starting from the cell surface. Arch. Microbiol., 167:325-31.
- Braun V, Mahren S, Ogierman M. 2003. Regulation of the Fecl-type ECF sigma factor by transmembrane signalling. Curr. Opin. Microbiol., 6:173-80.
- Brodolin K, Buckle M. 2001. Differential melting of the transcription start site associated with changes in RNA polymerase-promoter contacts in initiating transcription complexes. J. Mol. Biol., 307:25-30.
- Brodolin K, Zenkin N, Mustaev A, Mamaeva D, Heumann H. 2004. The σ^{70} subunit of RNA polymerase induces *lacUV5* promoter-proximal pausing of transcription. Nat. Struct. Mol. Biol., 11:551-7.
- Browning DF, Busby SJ. 2004. The regulation of bacterial transcription initiation. Nat. Rev. Microbiol., 2:57-65.
- Bruckner R, Titgemeyer F. 2002. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. FEMS Microbiol. Lett., 209:141-8.
- Buc H, McClure WR. 1985. Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the *lac UV5* promoter. Evidence for a sequential mechanism involving three steps. Biochemistry, 24:2712-23.
- Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD. 2000. The bacterial enhancer-dependent σ^{54} (σ^N) transcription factor. J. Bacteriol., 182:4129-36.
- Buckle M, Fritsch A, Roux P, Geiselmann J, Buc H. 1991a. Kinetic studies on promoter-RNA polymerase complexes. Methods Enzymol., 208:236-58.
- Buckle M, Geiselmann J, Kolb A, Buc H. 1991b. Protein-DNA cross-linking at the *lac* promoter. Nucleic Acids Res., 19:833-40.
- Buckle M, Pemberton IK, Jacquet MA, Buc H. 1999. The kinetics of sigma subunit directed promoter recognition by *E. coli* RNA polymerase. J. Mol. Biol., 285:955-64.

- Buhr A, Daniels GA, Erni B. 1992. The glucose transporter of *Escherichia coli*. Mutants with impaired translocation activity that retain phosphorylation activity. J. Biol. Chem., **267**:3847-51.
- Bullock WO, Fernandez JM, Short JM. 1987. XL1-blue: A high efficiency plasmid transforming *recA* *Escherichia coli* strain with β -galactosidase selection. Bio. Techniques, **5**:376.
- Burgess RR, Anthony L. 2001. How sigma docks to RNA polymerase and what sigma does. Curr. Opin. Microbiol., **4**:126-31.
- Burgess RR, Travers AA, Dunn JJ, Bautz EK. 1969. Factor stimulating transcription by RNA polymerase. Nature, **221**:43-6.
- Burns H, Minchin S. 1994. Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. Nucleic Acids Res., **22**:3840-5.
- Burns HD, Belyaeva TA, Busby SJ, Minchin SD. 1996. Temperature-dependence of open-complex formation at two *Escherichia coli* promoters with extended -10 sequences. Biochem. J., **317**:305-11.
- Burns HD, Ishihama A, Minchin SD. 1999. Open complex formation during transcription initiation at the *Escherichia coli* galP1 promoter: the role of the RNA polymerase α subunit at promoters lacking an UP-element. Nucleic Acids Res., **27**:2051-6.
- Burr T, Mitchell J, Kolb A, Minchin S, Busby S. 2000. DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. Nucleic Acids Res., **28**:1864-70.
- Busby S, Ebright RH. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell, **79**:743-6.
- Busby S, Ebright RH. 1997. Transcription activation at class II CAP-dependent promoters. Mol. Microbiol., **23**:853-9.
- Busby S, Ebright RH. 1999. Transcription activation by catabolite activator protein (CAP). J. Mol. Biol., **293**:199-213.
- Busby S, Spassky A, Chan B. 1987. RNA polymerase makes important contacts upstream from base pair -49 at the *Escherichia coli* galactose operon P1 promoter. Gene, **53**:145-52.
- C
- Cai M, Williams DC, Jr., Wang G, Lee BR, Peterkofsky A, Clore GM. 2003. Solution structure of the phosphoryl transfer complex between the signal-transducing protein IIA^{Glucose} and the cytoplasmic domain of the glucose transporter IICB^{Glucose} of the *Escherichia coli* glucose phosphotransferase system. J. Biol. Chem., **278**:25191-206.
- Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, Bryant DA, Darst SA, Cowburn D, Muir TW. 2002. Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. Proc. Natl. Acad. Sci. USA, **99**:8536-41.
- Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, Trester-Zedlitz ML, Darst SA. 2002. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol. Cell, **9**:527-39.
- Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, Gross CA, Darst SA. 2003. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. Mol. Cell, **11**:1067-78.
- Cannon W, Gallegos MT, Casaz P, Buck M. 1999. Amino-terminal sequences of σ^N (σ^{54}) inhibit RNA polymerase isomerization. Genes Dev., **13**:357-70.
- Carpousis AJ, Gralla JD. 1980. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation *in vitro* at the *lacUV5* promoter. Biochemistry, **19**:3245-53.
- Casaz P, Buck M. 1999. Region I modifies DNA-binding domain conformation of σ^{54} within the holoenzyme. J. Mol. Biol., **285**:507-14.
- Cashel M, Gentry DR, Hernandez VJ, Vinella D. 1996. The stringent response. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. Escherichia coli and Salmonella: cellular and molecular biology 2nd edn. Washington DC: American Society for microbiology press. pp 1458-1496.
- Chahla M, Wooll J, Laue TM, Nguyen N, Senear DF. 2003. Role of protein-protein bridging interactions on cooperative assembly of DNA-bound CRP-CytR-CRP complex and regulation of the *Escherichia coli* CytR regulon. Biochemistry, **42**:3812-25.
- Chan B, Busby S. 1989. Recognition of nucleotide sequences at the *Escherichia coli* galactose operon P1 promoter by RNA polymerase. Gene, **84**:227-36.
- Chan CL, Gross CA. 2001. The anti-initial transcribed sequence, a portable sequence that impedes promoter escape, requires σ^{70} for function. J. Biol. Chem., **276**:38201-9.
- Charpentier B, Bardey V, Robas N, Branlant C. 1998. The EII^{Glc} protein is involved in glucose-mediated activation of *Escherichia coli* *gapA* and *gapB-pgk* transcription. J. Bacteriol., **180**:6476-83.
- Charpentier B, Branlant C. 1994. The *Escherichia coli* *gapA* gene is transcribed by the vegetative RNA polymerase holoenzyme $E\sigma^{70}$ and by the heat shock RNA polymerase $E\sigma^{32}$. J. Bacteriol., **176**:830-9.
- Checroun C, Bordes P, Leroy O, Kolb A, Gutierrez C. 2004. Interactions between the 2.4 and 4.2

- regions of σ^S , the stress-specific sigma factor of *Escherichia coli*, and the -10 and -35 promoter elements. Nucleic Acids Res., 32:45-53. Print 2004.
- Chen H, Tang H, Ebricht RH. 2003. Functional interaction between RNA polymerase alpha subunit C-terminal domain and σ^{70} in UP-element- and activator-dependent transcription. Mol. Cell, 11:1621-33.
- Chen MH, Takeda S, Yamada H, Ishii Y, Yamashino T, Mizuno T. 2001. Characterization of the RcsC-->YojN-->RcsB phosphorelay signaling pathway involved in capsular synthesis in *Escherichia coli*. Biosci. Biotechnol. Biochem., 65:2364-7.
- Chen S, Gunasekera A, Zhang X, Kunkel TA, Ebricht RH, Berman HM. 2001b. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: alteration of DNA binding specificity through alteration of DNA kinking. J. Mol. Biol., 314:75-82.
- Chen S, Vojtechovsky J, Parkinson GN, Ebricht RH, Berman HM. 2001a. Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. J. Mol. Biol., 314:63-74.
- Cho J, Hamasaki K, Rando RR. 1998. The binding site of a specific aminoglycoside binding RNA molecule. Biochemistry, 37:4985-92.
- Choy HE, Adhya S. 1992. Control of *gal* transcription through DNA looping: inhibition of the initial transcribing complex. Proc. Natl. Acad. Sci. USA, 89:11264-8.
- Choy HE, Adhya S. 1996. Negative control. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. Escherichia coli and Salmonella: cellular and molecular biology 2nd edn. Washington DC: American Society for microbiology press. pp 1287-1299.
- Choy HE, Park SW, Aki T, Parrack P, Fujita N, Ishihama A, Adhya S. 1995. Repression and activation of transcription by Gal and Lac repressors: involvement of α subunit of RNA polymerase. EMBO J., 14:4523-9.
- Claret L, Hughes C. 2002. Interaction of the atypical prokaryotic transcription activator FlhD₂C₂ with early promoters of the flagellar gene hierarchy. J. Mol. Biol., 321:185-99.
- Cohen SN, Chang AC, Hsu L. 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. USA, 69:2110-4.
- Collis CM, Molloy PL, Both GW, Drew HR. 1989. Influence of the sequence-dependent flexure of DNA on transcription in *E. coli*. Nucleic Acids Res., 17:9447-68.
- Coulombe B, Burton ZF. 1999. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. Microbiol. Mol. Biol. Rev., 63:457-78.
- Craig ML, Suh WC, Record MT, Jr. 1995. HO. and DNase I probing of E σ^{70} RNA polymerase--lambda PR promoter open complexes: Mg²⁺ binding and its structural consequences at the transcription start site. Biochemistry, 34:15624-32.
- Craig ML, Tsodikov OV, McQuade KL, Schlx PE, Jr., Capp MW, Saecker RM, Record MT, Jr. 1998. DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. J. Mol. Biol., 283:741-56.
- Cramer P. 2002. Multisubunit RNA polymerases. Curr. Opin. Struct. Biol., 12:89-97.
- Crasnier M, Dumay V, Danchin A. 1994. The catalytic domain of *Escherichia coli* K-12 adenylate cyclase as revealed by deletion analysis of the *cya* gene. Mol. Gen. Genet., 243:409-16.
- Crothers DM, Haran TE, Nadeau JG. 1990. Intrinsically bent DNA. J. Biol. Chem., 265:7093-6.

D

- Dai X, Rothman-Denes LB. 1999. DNA structure and transcription. Curr. Opin. Microbiol., 2:126-30.
- Darst SA. 2001. Bacterial RNA polymerase. Curr. Opin. Struct. Biol., 11:155-62.
- Darst SA, Kubalek EW, Kornberg RD. 1989. Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. Nature, 340:730-2.
- Darst SA, Opalka N, Chacon P, Polyakov A, Richter C, Zhang G, Wriggers W. 2002. Conformational flexibility of bacterial RNA polymerase. Proc. Natl. Acad. Sci. USA, 99:4296-301.
- Darst SA, Polyakov A, Richter C, Zhang G. 1998. Insights into *Escherichia coli* RNA polymerase structure from a combination of x-ray and electron crystallography. J. Struct. Biol., 124:115-22.
- de Crombrughe B, Busby S, Buc H. 1984. Cyclic AMP receptor protein: role in transcription activation. Science, 224:831-8.
- De Reuse H, Danchin A. 1988. The *ptsH*, *ptsI*, and *crr* genes of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system: a complex operon with several modes of transcription. J. Bacteriol., 170:3827-37.
- De Reuse H, Kolb A, Danchin A. 1992. Positive regulation of the expression of the *Escherichia coli pts* operon. Identification of the regulatory regions. J. Mol. Biol., 226:623-35.

E

- Decker K, Plumbridge J, Boos W. 1998. Negative transcriptional regulation of a positive regulator: the expression of *maltT*, encoding the transcriptional activator of the maltose regulon of *Escherichia coli*, is negatively controlled by Mlc. Mol. Microbiol., 27:381-90.
- DeGrazia H, Harman JG, Tan GS, Wartell RM. 1990. Investigation of the cAMP receptor protein secondary structure by Raman spectroscopy. Biochemistry, 29:3557-62.
- DeHaseth PL, Helmann JD. 1995. Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol. Microbiol., 16:817-24.
- DeHaseth PL, Zupancic ML, Record MT, Jr. 1998. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. J. Bacteriol., 180:3019-25.
- Delihans N, Forst S. 2001. MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. J. Mol. Biol., 313:1-12.
- Della-Seta F, Boschi-Muller S, Vignais ML, Branlant G. 1997. Characterization of *Escherichia coli* strains with *gapA* and *gapB* genes deleted. J. Bacteriol., 179:5218-21.
- Deuschle U, Kammerer W, Gentz R, Bujard H. 1986. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. EMBO J., 5:2987-94.
- Dombroski AJ, Walter WA, Gross CA. 1993. Amino-terminal amino acids modulate σ -factor DNA-binding activity. Genes Dev., 7:2446-55.
- Dombroski AJ, Walter WA, Record MT, Jr., Siegel DA, Gross CA. 1992. Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. Cell, 70:501-12.
- Dorman CJ. 2004. H-NS: a universal regulator for a dynamic genome. Nat. Rev. Microbiol., 2:391-400.
- Dorman CJ, Deighan P. 2003. Regulation of gene expression by histone-like proteins in bacteria. Curr. Opin. Genet. Dev., 13:179-84.
- Drew HR, Weeks JR, Travers AA. 1985. Negative supercoiling induces spontaneous unwinding of a bacterial promoter. EMBO J., 4:1025-32.
- Drlica K. 1992. Control of bacterial DNA supercoiling. Mol. Microbiol., 6:425-33.
- Drlica K, Malik M, Rouvière-Yaniv J. 1992. Intracellular DNA Supercoiling in Bacteria. Nucleic Acids and Molecular Biology, 6:55-66.
- Duee E, Olivier-Deyris L, Fanchon E, Corbier C, Branlant G, Dideberg O. 1996. Comparison of the structures of wild-type and a N313T mutant of *Escherichia coli* glyceraldehyde 3-phosphate dehydrogenases: implication for NAD binding and cooperativity. J. Mol. Biol., 257:814-38.
- Ebright RH. 1993. Transcription activation at Class I CAP-dependent promoters. Mol. Microbiol., 8:797-802.
- Ebright RH. 2000. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J. Mol. Biol., 304:687-98.
- Ebright RH, Ebright YW, Gunasekera A. 1989. Consensus DNA site for the *Escherichia coli* catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the *E. coli lac* DNA site. Nucleic Acids Res., 17:10295-305.
- Ebright RH, Kolb A, Buc H, Kunkel TA, Krakow JS, Beckwith J. 1987. Role of glutamic acid-181 in DNA-sequence recognition by the catabolite gene activator protein (CAP) of *Escherichia coli*: altered DNA-sequence-recognition properties of [Val181]CAP and [Leu181]CAP. Proc. Natl. Acad. Sci. USA, 84:6083-7.
- Ederth J, Artsimovitch I, Isaksson LA, Landick R. 2002. The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. J. Biol. Chem., 277:37456-63.
- Egger LA, Park H, Inouye M. 1997. Signal transduction via the histidyl-aspartyl phosphorelay. Genes Cells, 2:167-84.
- Eichenberger P, Dethiollaz S, Buc H, Geiselman J. 1997. Structural kinetics of transcription activation at the *maltT* promoter of *Escherichia coli* by UV laser footprinting. Proc. Natl. Acad. Sci. USA, 94:9022-7.
- Eikmanns BJ. 1992. Identification, sequence analysis, and expression of a *Corynebacterium glutamicum* gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomerase. J. Bacteriol., 174:6076-86.
- Ellinger T, Behnke D, Bujard H, Gralla JD. 1994a. Stalling of *Escherichia coli* RNA polymerase in the +6 to +12 region *in vivo* is associated with tight binding to consensus promoter elements. J. Mol. Biol., 239:455-65.
- Ellinger T, Behnke D, Knaus R, Bujard H, Gralla JD. 1994b. Context-dependent effects of upstream A-tracts. Stimulation or inhibition of *Escherichia coli* promoter function. J. Mol. Biol., 239:466-75.
- Eschenlauer AC, Reznikoff WS. 1991. *Escherichia coli* catabolite gene activator protein mutants defective in positive control of *lac* operon transcription. J. Bacteriol., 173:5024-9.
- Estrem ST, Gaal T, Ross W, Gourse RL. 1998. Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. USA, 95:9761-6.

Estrem ST, Ross W, Gaal T, Chen ZW, Niu W, Ebright RH, Gourse RL. 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. Genes Dev., 13:2134-47.

F

- Falke JJ, Hazelbauer GL. 2001. Transmembrane signaling in bacterial chemoreceptors. Trends Biochem. Sci., 26:257-65.
- Fenton MS, Gralla JD. 2001. Function of the bacterial TATAAT -10 element as single-stranded DNA during RNA polymerase isomerization. Proc. Natl. Acad. Sci. USA, 98:9020-5.
- Fenton MS, Lee SJ, Gralla JD. 2000. *Escherichia coli* promoter opening and -10 recognition: mutational analysis of σ^{70} . EMBO J., 19:1130-7.
- Ferguson AD, Chakraborty R, Smith BS, Esser L, van der Helm D, Deisenhofer J. 2002. Structural basis of gating by the outer membrane transporter FecA. Science, 295:1715-9.
- Figueroa-Bossi N, Guerin M, Rahmouni R, Leng M, Bossi L. 1998. The supercoiling sensitivity of a bacterial tRNA promoter parallels its responsiveness to stringent control. EMBO J., 17:2359-67.
- Fillinger S, Boschi-Muller S, Azza S, Dervyn E, Branlant G, Aymerich S. 2000. Two glyceraldehyde-3-phosphate dehydrogenases with opposite physiological roles in a nonphotosynthetic bacterium. J. Biol. Chem., 275:14031-7.
- Finn RD, Orlova EV, Gowen B, Buck M, van Heel M. 2000. *Escherichia coli* RNA polymerase core and holoenzyme structures. EMBO J., 19:6833-44.
- Forst S, Delgado J, Inouye M. 1989. Phosphorylation of OmpR by the osmosensor EnvZ modulates expression of the *ompF* and *ompC* genes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 86:6052-6.
- Fox DK, Presper KA, Adhya S, Roseman S, Garges S. 1992. Evidence for two promoters upstream of the *pts* operon: regulation by the cAMP receptor protein regulatory complex. Proc. Natl. Acad. Sci. USA, 89:7056-9.

G

- Gaal T, Bartlett MS, Ross W, Turnbough CL, Jr., Gourse RL. 1997. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. Science, 278:2092-7.
- Gaal T, Gourse RL. 1990. Guanosine 3'-diphosphate 5'-diphosphate is not required for growth rate-dependent control of rRNA synthesis in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 87:5533-7.
- Galas DJ, Eggert M, Waterman MS. 1985. Rigorous pattern-recognition methods for DNA sequences. Analysis of promoter sequences from *Escherichia coli*. J. Mol. Biol., 186:117-28.
- Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rudiger S, Schonfeld HJ, Schirra C, Bujard H, Bukau B. 1996. A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor σ^{32} . EMBO J., 15:607-17.
- Garcia-Alles LF, Navdaeva V, Haenni S, Erni B. 2002a. The glucose-specific carrier of the *Escherichia coli* phosphotransferase system. Eur. J. Biochem., 269:4969-80.
- Garcia-Alles LF, Zahn A, Erni B. 2002b. Sugar recognition by the glucose and mannose permeases of *Escherichia coli*. Steady-state kinetics and inhibition studies. Biochemistry, 41:10077-86.
- Gardella T, Moyle H, Susskind MM. 1989. A mutant *Escherichia coli* σ^{70} subunit of RNA polymerase with altered promoter specificity. J. Mol. Biol., 206:579-90.
- Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM. 1999. Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. Nat. Struct. Biol., 6:166-73.
- Gartenberg MR, Crothers DM. 1988. DNA sequence determinants of CAP-induced bending and protein binding affinity. Nature, 333:824-9.
- Gaston K, Bell A, Kolb A, Buc H, Busby S. 1990. Stringent spacing requirements for transcription activation by CRP. Cell, 62:733-43.
- Gaston K, Kolb A, Busby S. 1989. Binding of the *Escherichia coli* cyclic AMP receptor protein to DNA fragments containing consensus nucleotide sequences. Biochem. J., 261:649-53.
- Gausing K. 1977. Regulation of ribosome production in *Escherichia coli*: synthesis and stability of ribosomal RNA and of ribosomal protein messenger RNA at different growth rates. J. Mol. Biol., 115:335-54.
- Geissmann TA, Touati D. 2004. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. EMBO J., 23:396-405.
- Georgellis D, Barlow T, Arvidson S, von Gabain A. 1993. Retarded RNA turnover in *Escherichia coli*: a means of maintaining gene expression during anaerobiosis. Mol. Microbiol., 9:375-81.
- Georgellis D, Kwon O, Lin EC. 2001. Quinones as the redox signal for the arc two-component system of bacteria. Science, 292:2314-6.
- Ghosh P, Ishihama A, Chatterji D. 2001. *Escherichia coli* RNA polymerase subunit omega and its N-terminal domain bind full-length β' to facilitate

- incorporation into the $\alpha 2\beta$ subassembly. Eur. J. Biochem., 268:4621-7.
- Ghosh P, Ramakrishnan C, Chatterji D. 2003. Inter-subunit recognition and manifestation of segmental mobility in *Escherichia coli* RNA polymerase: a case study with ω - β' interaction. Biophys. Chem., 103:223-37.
- Giladi H, Koby S, Prag G, Engelhorn M, Geiselman J, Oppenheim AB. 1998. Participation of IHF and a distant UP element in the stimulation of the phage lambda PL promoter. Mol. Microbiol., 30:443-51.
- Gillen KL, Hughes KT. 1991. Negative regulatory loci coupling flagellin synthesis to flagellar assembly in *Salmonella typhimurium*. J. Bacteriol., 173:2301-10.
- Ginsburg A, Peterkofsky A. 2002. Enzyme I: the gateway to the bacterial phosphoenolpyruvate:sugar phosphotransferase system. Arch. Biochem. Biophys., 397:273-8.
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science, 292:1876-82.
- Gonzalez-Gil G, Kahmann R, Muskhelishvili G. 1998. Regulation of *crp* transcription by oscillation between distinct nucleoprotein complexes. EMBO J., 17:2877-85.
- Goosen N, van de Putte P. 1995. The regulation of transcription initiation by integration host factor. Mol. Microbiol., 16:1-7.
- Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MH, Jr. 2004. Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. J. Bacteriol., 186:3516-24.
- Gottesman S. 2004. The Small RNA Regulators of *Escherichia coli*: Roles and Mechanisms. Annu. Rev. Microbiol., 16:16.
- Gourse RL, Ross W, Gaal T. 2000. UPs and downs in bacterial transcription initiation: the role of the α subunit of RNA polymerase in promoter recognition. Mol. Microbiol., 37:687-95.
- Grimaud R, Ezraty B, Mitchell JK, Lafitte D, Briand C, Derrick PJ, Barras F. 2001. Repair of oxidized proteins. Identification of a new methionine sulfoxide reductase. J. Biol. Chem., 276:48915-20.
- Gruber TM, Gross CA. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol., 57:441-66.
- Gruber TM, Markov D, Sharp MM, Young BA, Lu CZ, Zhong HJ, Artsimovitch I, Geszvain KM, Arthur TM, Burgess RR, Landick R, Severinov K, Gross CA. 2001. Binding of the initiation factor σ^{70} to core RNA polymerase is a multistep process. Mol. Cell, 8:21-31.
- Gschaedler A, Robas N, Boudrant J, Branlant C. 1999. Effects of pulse addition of carbon sources on continuous cultivation of *Escherichia coli* containing recombinant *E. coli gapA* gene. Biotech. Bioeng., 63:712-720.
- Gudapaty S, Suzuki K, Wang X, Babitzke P, Romeo T. 2001. Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*. J. Bacteriol., 183:6017-27.
- Guo Y, Gralla JD. 1998. Promoter opening via a DNA fork junction binding activity. Proc. Natl. Acad. Sci. USA, 95:11655-60.

H

- Hall MN, Silhavy TJ. 1981. Genetic analysis of the major outer membrane proteins of *Escherichia coli*. Annu. Rev. Genet., 15:91-142.
- Hanamura A, Aiba H. 1991. Molecular mechanism of negative autoregulation of *Escherichia coli crp* gene. Nucleic Acids Res., 19:4413-9.
- Hanamura A, Aiba H. 1992. A new aspect of transcriptional control of the *Escherichia coli crp* gene: positive autoregulation. Mol. Microbiol., 6:2489-97.
- Hantke K. 2001. Iron and metal regulation in bacteria. Curr. Opin. Microbiol., 4:172-7.
- Harley CB, Reynolds RP. 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res., 15:2343-61.
- Harman JG. 2001. Allosteric regulation of the cAMP receptor protein. Biochim. Biophys. Acta, 1547:1-17.
- Harris JL, Waters M. 1976. Glyceraldehyde 3-phosphate dehydrogenase. Academic Press. New York, N. Y.
- Hartmuth K, Urlaub H, Vornlocher HP, Will CL, Gentzel M, Wilm M, Luhrmann R. 2002. Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc. Natl. Acad. Sci. USA, 99:16719-24.
- Hatfield GW, Benham CJ. 2002. DNA topology-mediated control of global gene expression in *Escherichia coli*. Annu. Rev. Genet., 36:175-203.
- Hayward RS, Igarashi K, Ishihama A. 1991. Functional specialization within the α -subunit of *Escherichia coli* RNA polymerase. J. Mol. Biol., 221:23-9.
- Head CG, Tardy A, Kenney LJ. 1998. Relative binding affinities of OmpR and OmpR-phosphate at the *ompF* and *ompC* regulatory sites. J. Mol. Biol., 281:857-70.
- Helmann JD. 1995. Compilation and analysis of *Bacillus subtilis* σ^A -dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res., 23:2351-60.
- Helmann JD. 2002. The extracytoplasmic function (ECF) sigma factors. Adv. Microb. Physiol., 46:47-110.

- Helmann JD, Chamberlin MJ. 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem., 57:839-72.
- Helmann JD, deHaseth PL. 1999. Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. Biochemistry, 38:5959-67.
- Hengen PN, Bartram SL, Stewart LE, Schneider TD. 1997. Information analysis of Fis binding sites. Nucleic Acids Res., 25:4994-5002.
- Hengge-Aronis R. 1996. Back to log phase: σ^S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. Mol. Microbiol., 21:887-93.
- Hengge-Aronis R. 2002. Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. Microbiol. Mol. Biol. Rev., 66:373-95.
- Heyde M, Laloi P, Portalier R. 2000. Involvement of carbon source and acetyl phosphate in the external-pH-dependent expression of porin genes in *Escherichia coli*. J. Bacteriol., 182:198-202.
- Heyduk T, Heyduk E, Severinov K, Tang H, Ebright RH. 1996. Determinants of RNA polymerase alpha subunit for interaction with β , β' , and σ subunits: hydroxyl-radical protein footprinting. Proc. Natl. Acad. Sci. USA, 93:10162-6.
- Heyduk T, Lee JC. 1989. *Escherichia coli* cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. Biochemistry, 28:6914-24.
- Heyduk T, Lee JC. 1990. Application of fluorescence energy transfer and polarization to monitor *Escherichia coli* cAMP receptor protein and *lac* promoter interaction. Proc. Natl. Acad. Sci. USA, 87:1744-8.
- Hirvonen CA, Ross W, Wozniak CE, Marasco E, Anthony JR, Aiyar SE, Newburn VH, Gourse RL. 2001. Contributions of UP elements and the transcription factor FIS to expression from the seven *rrn* P1 promoters in *Escherichia coli*. J. Bacteriol., 183:6305-14.
- Hoch JA. 2000. Two-component and phosphorelay signal transduction. Curr. Opin. Microbiol., 3:165-70.
- Hockensmith JW, Kubasek WL, Vorachek WR, von Hippel PH. 1986. Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. J. Biol. Chem., 261:3512-8.
- Hogema BM, Arents JC, Bader R, Eijkemans K, Inada T, Aiba H, Postma PW. 1998b. Inducer exclusion by glucose 6-phosphate in *Escherichia coli*. Mol. Microbiol., 28:755-65.
- Hogema BM, Arents JC, Bader R, Eijkemans K, Yoshida H, Takahashi H, Aiba H, Postma PW. 1998a. Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIA^{Glc}. Mol. Microbiol., 30:487-98.
- Hogema BM, Arents JC, Bader R, Postma PW. 1999. Autoregulation of lactose uptake through the LacY permease by enzyme IIA^{Glc} of the PTS in *Escherichia coli* K-12. Mol. Microbiol., 31:1825-33.
- Hogema BM, Arents JC, Inada T, Aiba H, van Dam K, Postma PW. 1997. Catabolite repression by glucose 6-phosphate, gluconate and lactose in *Escherichia coli*. Mol. Microbiol., 24:857-67.
- Hosono K, Kakuda H, Ichihara S. 1995. Decreasing accumulation of acetate in a rich medium by *Escherichia coli* on introduction of genes on a multicopy plasmid. Biosci. Biotechnol. Biochem., 59:256-61.
- Hsieh LS, Rouviere-Yaniv J, Drlica K. 1991. Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. J. Bacteriol., 173:3914-7.
- Hsu LM. 2002a. Open season on RNA polymerase. Nat. Struct. Biol., 9:502-4.
- Hsu LM. 2002b. Promoter clearance and escape in prokaryotes. Biochim. Biophys. Acta, 1577:191-207.
- Hsu LM, Giannini JK, Leung TW, Crosthwaite JC. 1991. Upstream sequence activation of *Escherichia coli argT* promoter *in vivo* and *in vitro*. Biochemistry, 30:813-22.
- Hsu LM, Vo NV, Chamberlin MJ. 1995. *Escherichia coli* transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression *in vivo* and *in vitro*. Proc. Natl. Acad. Sci. USA, 92:11588-92.
- Hsu LM, Vo NV, Kane CM, Chamberlin MJ. 2003. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. I. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. Biochemistry, 42:3777-86.
- Huang KJ, Igo MM. 1996. Identification of the bases in the *ompF* regulatory region, which interact with the transcription factor OmpR. J. Mol. Biol., 262:615-28.
- Huang L, Tsui P, Freundlich M. 1992. Positive and negative control of *ompB* transcription in *Escherichia coli* by cyclic AMP and the cyclic AMP receptor protein. J. Bacteriol., 174:664-70.
- Hughes KT, Mathee K. 1998. The anti-sigma factors. Annu. Rev. Microbiol., 52:231-86.
- Igarashi K, Fujita N, Ishihama A. 1991a. Identification of a subunit assembly domain in the α subunit of *Escherichia coli* RNA polymerase. J. Mol. Biol., 218:1-6.
- Igarashi K, Ishihama A. 1991c. Bipartite functional map of the *E. coli* RNA polymerase alpha subunit: involvement of the C-terminal region in

- transcription activation by cAMP-CRP. Cell, **65**:1015-22.
- Iglesias AA, Losada M. 1988. Purification and kinetic and structural properties of spinach leaf NADP-dependent nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase. Arch. Biochem. Biophys., **260**:830-40.
- Imlay JA. 2003. Pathways of oxidative damage. Annu. Rev. Microbiol., **57**:395-418.
- Inada T, Kimata K, Aiba H. 1996a. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. Genes Cells, **1**:293-301.
- Ishige K, Nagasawa S, Tokishita S, Mizuno T. 1994. A novel device of bacterial signal transducers. EMBO J., **13**:5195-202.
- Ishihama A. 1981. Subunit of assembly of *Escherichia coli* RNA polymerase. Adv. Biophys., **14**:1-35.
- Ishihama A. 2000. Functional modulation of *Escherichia coli* RNA polymerase. Annu. Rev. Microbiol., **54**:499-518.
- Ishizuka H, Hanamura A, Inada T, Aiba H. 1994. Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene. EMBO J., **13**:3077-82.
- Ishizuka H, Hanamura A, Kunimura T, Aiba H. 1993. A lowered concentration of cAMP receptor protein caused by glucose is an important determinant for catabolite repression in *Escherichia coli*. Mol. Microbiol., **10**:341-50.
- J
- Jeon YH, Yamazaki T, Otomo T, Ishihama A, Kyogoku Y. 1997. Flexible linker in the RNA polymerase α subunit facilitates the independent motion of the C-terminal activator contact domain. J. Mol. Biol., **267**:953-62.
- Jeong JY, Kim YJ, Cho N, Shin D, Nam TW, Ryu S, Seok YJ. 2004. Expression of *ptsG* encoding the major glucose transporter is regulated by ArcA in *Escherichia coli*. J. Biol. Chem., **13**:13.
- Jishage M, Dasgupta D, Ishihama A. 2001. Mapping of the Rsd contact site on the σ^{70} subunit of *Escherichia coli* RNA polymerase. J. Bacteriol., **183**:2952-6.
- Jishage M, Ishihama A. 1995. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of σ^{70} and σ^{38} . J. Bacteriol., **177**:6832-5.
- Jishage M, Ishihama A. 1998. A stationary phase protein in *Escherichia coli* with binding activity to the major σ subunit of RNA polymerase. Proc. Natl. Acad. Sci. USA, **95**:4953-8.
- Jishage M, Ishihama A. 1999. Transcriptional organization and *in vivo* role of the *Escherichia coli* *rsd* gene, encoding the regulator of RNA polymerase sigma D. J. Bacteriol., **181**:3768-76.
- Jishage M, Iwata A, Ueda S, Ishihama A. 1996. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. J. Bacteriol., **178**:5447-51.
- Jores L, Wagner R. 2003. Essential steps in the ppGpp-dependent regulation of bacterial ribosomal RNA promoters can be explained by substrate competition. J. Biol. Chem., **278**:16834-43.
- Josaitis CA, Gaal T, Gourse RL. 1995. Stringent control and growth-rate-dependent control have nonidentical promoter sequence requirements. Proc. Natl. Acad. Sci. USA, **92**:1117-21.
- Juang YL, Helmann JD. 1994. A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids. J. Mol. Biol., **235**:1470-88.
- K
- Kahn JD, Crothers DM. 1998. Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics. J. Mol. Biol., **276**:287-309.
- Kammerer W, Deuschle U, Gentz R, Bujard H. 1986. Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. EMBO J., **5**:2995-3000.
- Kanehara K, Ito K, Akiyama Y. 2002. YaeL (EcfE) activates the σ^E pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. Genes Dev., **16**:2147-55.
- Kapanidis AN, Ebright YW, Ludescher RD, Chan S, Ebright RH. 2001. Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution. J. Mol. Biol., **312**:453-68.
- Katayama A, Fujita N, Ishihama A. 2000. Mapping of subunit-subunit contact surfaces on the β' subunit of *Escherichia coli* RNA polymerase. J. Biol. Chem., **275**:3583-92.
- Kauffmann B, Favier F, Olry A, Boschi-Muller S, Carpentier P, Branlant G, Aubry A. 2002. Crystallization and preliminary X-ray diffraction studies of the peptide methionine sulfoxide reductase B domain of *Neisseria meningitidis* PILB. Acta Crystallogr. D Biol. Crystallogr., **58**:1467-9.
- Keener J, Kustu S. 1988. Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. Proc. Natl. Acad. Sci. USA, **85**:4976-80.
- Keener J, Nomura M. 1993. Dominant lethal phenotype of a mutation in the -35 recognition region of *Escherichia coli* σ^{70} . Proc. Natl. Acad. Sci. USA, **90**:1751-5.

- Keener J, Nomura M. 1996. Regulation of ribosome synthesis. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. *Escherichia coli and Salmonella: cellular and molecular biology 2nd edn*. Washington DC: American Society for microbiology press. pp 1417-1431.
- Keilty S, Rosenberg M. 1987. Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J. Biol. Chem.*, 262:6389-95.
- Kenney LJ. 2002. Structure/function relationships in OmpR and other winged-helix transcription factors. *Curr. Opin. Microbiol.*, 5:135-41.
- Kim I, Stiefel A, Plantor S, Angerer A, Braun V. 1997. Transcription induction of the ferric citrate transport genes via the N-terminus of the FecA outer membrane protein, the Ton system and the electrochemical potential of the cytoplasmic membrane. *Mol. Microbiol.*, 23:333-44.
- Kim SK, Makino K, Amemura M, Nakata A, Shinagawa H. 1995. Mutational analysis of the role of the first helix of region 4.2 of the σ^{70} subunit of *Escherichia coli* RNA polymerase in transcriptional activation by activator protein PhoB. *Mol. Gen. Genet.*, 248:1-8.
- Kim SY, Nam TW, Shin D, Koo BM, Seok YJ, Ryu S. 1999. Purification of Mlc and analysis of its effects on the *pts* expression in *Escherichia coli*. *J. Biol. Chem.*, 274:25398-402.
- Kimata K, Inada T, Tagami H, Aiba H. 1998. A global repressor (Mlc) is involved in glucose induction of the *ptsG* gene encoding major glucose transporter in *Escherichia coli*. *Mol. Microbiol.*, 29:1509-19.
- Kimata K, Takahashi H, Inada T, Postma P, Aiba H. 1997. cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 94:12914-9.
- Kimata K, Tanaka Y, Inada T, Aiba H. 2001. Expression of the glucose transporter gene, *ptsG*, is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J.*, 20:3587-95.
- Knaus R, Bujard H. 1988. PL of coliphage lambda: an alternative solution for an efficient promoter. *EMBO J.*, 7:2919-23.
- Koch C, Vandekerckhove J, Kahmann R. 1988. *Escherichia coli* host factor for site-specific DNA inversion: cloning and characterization of the *fis* gene. *Proc. Natl. Acad. Sci. USA*, 85:4237-41.
- Koebnik R, Locher KP, Van Gelder P. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.*, 37:239-53.
- Kojima S, Blair DF. 2004. The bacterial flagellar motor: structure and function of a complex molecular machine. *Int. Rev. Cytol.*, 233:93-134.
- Koksharova O, Schubert M, Shestakov S, Cerff R. 1998. Genetic and biochemical evidence for distinct key functions of two highly divergent GAPDH genes in catabolic and anabolic carbon flow of the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant. Mol. Biol.*, 36:183-94.
- Kolasa IK, Lozinski T, Wierzychowski KL. 2002. Effect of A(n) tracts within the UP element proximal subsite of a model promoter on kinetics of open complex formation by *Escherichia coli* RNA polymerase. *Acta Biochim. Pol.*, 49:659-69.
- Kolasa IK, Lozinski T, Wierzychowski KL. 2003. Effects of distortions by A-tracts of promoter B-DNA spacer region on the kinetics of open complex formation by *Escherichia coli* RNA polymerase. *Acta Biochim. Pol.*, 50:909-20.
- Kolb A, Busby S, Buc H, Garges S, Adhya S. 1993a. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.*, 62:749-95.
- Koo BM, Yoon MJ, Lee CR, Nam TW, Choe YJ, Jaffe H, Peterkofsky A, Seok YJ. 2004. A novel fermentation/respiration switch protein regulated by enzyme IIA^{Glc} in *Escherichia coli*. *J. Biol. Chem.*, 279:31613-21.
- Koo HS, Crothers DM. 1988. Calibration of DNA curvature and a unified description of sequence-directed bending. *Proc. Natl. Acad. Sci. USA*, 85:1763-7.
- Koo HS, Drak J, Rice JA, Crothers DM. 1990. Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry*, 29:4227-34.
- Koo HS, Wu HM, Crothers DM. 1986. DNA bending at adenine . thymine tracts. *Nature*, 320:501-6.
- Kornberg HL, Lambourne LT, Sproul AA. 2000. Facilitated diffusion of fructose via the phosphoenolpyruvate/glucose phosphotransferase system of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 97:1808-12.
- Korteland J, Tommassen J, Lugtenberg B. 1982. PhoE protein pore of the outer membrane of *Escherichia coli* K12 is a particularly efficient channel for organic and inorganic phosphate. *Biochim. Biophys. Acta*, 690:282-9.
- Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A, Darst SA. 2000. A structural model of transcription elongation. *Science*, 289:619-25.
- Kovacik RT. 1987. The 0 degree C closed complexes between *Escherichia coli* RNA polymerase and two promoters, T7-A3 and *lacUV5*. *J. Biol. Chem.*, 262:13654-61.
- Krin E, Sismeiro O, Danchin A, Bertin PN. 2002. The regulation of Enzyme IIA^{Glc} expression controls adenylate cyclase activity in *Escherichia coli*. *Microbiology*, 148:1553-9.

- Kubori T, Shimamoto N. 1996. A branched pathway in the early stage of transcription by *Escherichia coli* RNA polymerase. J. Mol. Biol., 256:449-57.
- Kumar A, Malloch RA, Fujita N, Smillie DA, Ishihama A, Hayward RS. 1993. The minus 35-recognition region of *Escherichia coli* σ^{70} is inessential for initiation of transcription at an "extended minus 10" promoter. J. Mol. Biol., 232:406-18.
- Kundu TK, Kusano S, Ishihama A. 1997. Promoter selectivity of *Escherichia coli* RNA polymerase sigmaF holoenzyme involved in transcription of flagellar and chemotaxis genes. J. Bacteriol., 179:4264-9.
- Kuo JT, Chang YJ, Tseng CP. 2003. Growth rate regulation of *lac* operon expression in *Escherichia coli* is cyclic AMP dependent. FEBS Lett., 553:397-402.
- Kuroda M, de Waard S, Mizushima K, Tsuda M, Postma P, Tsuchiya T. 1992. Resistance of the melibiose carrier to inhibition by the phosphotransferase system due to substitutions of amino acid residues in the carrier of *Salmonella typhimurium*. J. Biol. Chem., 267:18336-41.
- Kutsukake K. 1994. Excretion of the anti-sigma factor through a flagellar substructure couples flagellar gene expression with flagellar assembly in *Salmonella typhimurium*. Mol. Gen. Genet., 243:605-12.
- Kuznedelov K, Korzheva N, Mustaev A, Severinov K. 2002a. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. EMBO J., 21:1369-78.
- Kuznedelov K, Minakhin L, Niedziela-Majka A, Dove SL, Rogulja D, Nickels BE, Hochschild A, Heyduk T, Severinov K. 2002b. A role for interaction of the RNA polymerase flap domain with the σ subunit in promoter recognition. Science, 295:855-7.
- Landick R. 2004. Active-site dynamics in RNA polymerases. Cell, 116:351-3.
- Lange R, Hengge-Aronis R. 1991. Growth phase-regulated expression of *bolA* and morphology of stationary-phase *Escherichia coli* cells are controlled by the novel sigma factor σ^S . J. Bacteriol., 173:4474-81.
- Lange R, Hengge-Aronis R. 1994. The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. Genes Dev., 8:1600-12.
- Lanz R, Erni B. 1998. The glucose transporter of the *Escherichia coli* phosphotransferase system. Mutant analysis of the invariant arginines, histidines, and domain linker. J. Biol. Chem., 273:12239-43.
- Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebricht RH. 2004. Catabolite activator protein: DNA binding and transcription activation. Curr. Opin. Struct. Biol., 14:10-20.
- Lee IS, Lin J, Hall HK, Bearson B, Foster JW. 1995. The stationary-phase sigma factor σ^S (RpoS) is required for a sustained acid tolerance response in virulent *Salmonella typhimurium*. Mol. Microbiol., 17:155-67.
- Lee K, Zhan X, Gao J, Qiu J, Feng Y, Meganathan R, Cohen SN, Georgiou G. 2003. RraA, a protein inhibitor of RNase E activity that globally modulates RNA abundance in *E. coli*. Cell, 114:623-34.
- Lee SJ, Boos W, Bouche JP, Plumbridge J. 2000. Signal transduction between a membrane-bound transporter, PtsG, and a soluble transcription factor, Mlc, of *Escherichia coli*. EMBO J., 19:5353-61.
- Lee SJ, Gralla JD. 2001. σ^{38} (rpoS) RNA polymerase promoter engagement via -10 region nucleotides. J. Biol. Chem., 276:30064-71.
- Lee SJ, Gralla JD. 2002. Promoter use by σ^{38} (rpoS) RNA polymerase. Amino acid clusters for DNA binding and isomerization. J. Biol. Chem., 277:47420-7.
- Lesley SA, Burgess RR. 1989. Characterization of the *Escherichia coli* transcription factor σ^{70} : localization of a region involved in the interaction with core RNA polymerase. Biochemistry, 28:7728-34.
- Levy S, Zeng GQ, Danchin A. 1990. Cyclic AMP synthesis in *Escherichia coli* strains bearing known deletions in the *pts* phosphotransferase operon. Gene, 86:27-33.
- Li M, Moyle H, Susskind MM. 1994. Target of the transcriptional activation function of phage lambda cI protein. Science, 263:75-7.
- Lim HM, Lewis DE, Lee HJ, Liu M, Adhya S. 2003. Effect of varying the supercoiling of DNA on transcription and its regulation. Biochemistry, 42:10718-25.
- Lisser S, Margalit H. 1993. Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res., 21:1507-16.
- Liu M, Gupte G, Roy S, Bandwar RP, Patel SS, Garges S. 2003. Kinetics of transcription initiation at *lacP1*. Multiple roles of cyclic AMP receptor protein. J. Biol. Chem., 278:39755-61.
- Liu X, Ferenci T. 1998. Regulation of porin-mediated outer membrane permeability by nutrient limitation in *Escherichia coli*. J. Bacteriol., 180:3917-22.
- Liu X, Ferenci T. 2001. An analysis of multifactorial influences on the transcriptional control of *ompF* and *ompC* porin expression under nutrient limitation. Microbiology, 147:2981-9.

- Liu X, Matsumura P. 1994. The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons. J. Bacteriol., 176:7345-51.
- Lonetto M, Gribskov M, Gross CA. 1992. The σ^{70} family: sequence conservation and evolutionary relationships. J. Bacteriol., 174:3843-9.
- Lonetto MA, Rhodius V, Lamberg K, Kiley P, Busby S, Gross C. 1998. Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase σ^{70} subunit. J. Mol. Biol., 284:1353-65.
- Lozinski T, Adrych-Rozek K, Markiewicz WT, Wierchowski K. 1991. Effect of DNA bending in various regions of a consensus-like *Escherichia coli* promoter on its strength *in vivo* and structure of the open complex *in vitro*. Nucleic Acids Res., 19:2947-53.
- Lozinski T, Markiewicz WT, Wyrzykiewicz TK, Wierchowski KL. 1989. Effect of the sequence-dependent structure of the 17 bp AT spacer on the strength of consensus like *E. coli* promoters *in vivo*. Nucleic Acids Res., 17:3855-63.
- Lozinski T, Wierchowski KL. 1996. Effect of reversed orientation and length of An.Tn DNA bending sequences in the -35 and spacer domains of a consensus-like *Escherichia coli* promoter on its strength *in vivo* and gross structure of the open complex *in vitro*. Acta Biochim. Pol., 43:265-79.
- Lupas A, Stock J. 1989. Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. J. Biol. Chem., 264:17337-42.
- Lux R, Munasinghe VR, Castellano F, Lengeler JW, Corrie JE, Khan S. 1999. Elucidation of a PTS-carbohydrate chemotactic signal pathway in *Escherichia coli* using a time-resolved behavioral assay. Mol. Biol. Cell., 10:1133-46.
- M
- MacDonald D, Herbert K, Zhang X, Pologruto T, Lu P, Polgruto T. 2001. Solution structure of an A-tract DNA bend. J. Mol. Biol., 306:1081-98.
- Macnab RM. 1996. Flagella and Motility. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edn. Washington DC: American Society for microbiology press. pp 123-145.
- Madan Babu M, Teichmann SA. 2003a. Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. Trends Genet., 19:75-9.
- Madan Babu M, Teichmann SA. 2003b. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. Nucleic Acids Res., 31:1234-44.
- Maeda H, Fujita N, Ishihama A. 2000a. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res., 28:3497-503.
- Maeda H, Jishage M, Nomura T, Fujita N, Ishihama A. 2000b. Two extracytoplasmic function sigma subunits, σ^E and σ^{FecI} , of *Escherichia coli*: promoter selectivity and intracellular levels. J. Bacteriol., 182:1181-4.
- Mager WH, De Kruijff AJ. 1995. Stress-induced transcriptional activation. Microbiol. Rev., 59:506-31.
- Malhotra A, Severinova E, Darst SA. 1996. Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. Cell, 87:127-36.
- Mallik P, Pratt TS, Beach MB, Bradley MD, Undamatla J, Osuna R. 2004. Growth phase-dependent regulation and stringent control of *fis* are conserved processes in enteric bacteria and involve a single promoter (*fis P*) in *Escherichia coli*. J. Bacteriol., 186:122-35.
- Manson MD, Armitage JP, Hoch JA, Macnab RM. 1998. Bacterial locomotion and signal transduction. J. Bacteriol., 180:1009-22.
- Marr MT, Roberts JW. 1997. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science, 276:1258-60.
- Martin W, Brinkmann H, Savonna C, Cerff R. 1993. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci. USA, 90:8692-6.
- Martinez-Antonio A, Collado-Vides J. 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. Curr. Opin. Microbiol., 6:482-9.
- Masse E, Escorcía FE, Gottesman S. 2003a. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes Dev., 17:2374-83.
- Masse E, Gottesman S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 99:4620-5.
- Masse E, Majdalani N, Gottesman S. 2003b. Regulatory roles for small RNAs in bacteria. Curr. Opin. Microbiol., 6:120-4.
- Matsubara M, Kitaoka SI, Takeda SI, Mizuno T. 2000. Tuning of the porin expression under anaerobic growth conditions by his-to-Asp cross-phosphorelay through both the EnvZ-osmosensor and ArcB-anaerosensor in *Escherichia coli*. Genes Cells, 5:555-69.
- McAllister CF, Achberger EC. 1989. Rotational orientation of upstream curved DNA affects promoter function in *Bacillus subtilis*. J. Biol. Chem., 264:10451-6.
- McCleary WR, Stock JB, Ninfa AJ. 1993. Is acetyl phosphate a global signal in *Escherichia coli*? J. Bacteriol., 175:2793-8.

- McLeod SM, Aiyar SE, Gourse RL, Johnson RC. 2002. The C-terminal domains of the RNA polymerase α subunits: contact site with Fis and localization during co-activation with CRP at the *Escherichia coli proP* P2 promoter. J. Mol. Biol., 316:517-29.
- McLeod SM, Johnson RC. 2001. Control of transcription by nucleoid proteins. Curr. Opin. Microbiol., 4:152-9.
- Mecenas J, Cowing DW, Gross CA. 1991. Development of RNA polymerase-promoter contacts during open complex formation. J. Mol. Biol., 220:585-97.
- Mekler V, Kortkhonjia E, Mukhopadhyay J, Knight J, Revyakin A, Kapanidis AN, Niu W, Ebright YW, Levy R, Ebright RH. 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell, 108:599-614.
- Meng W, Belyaeva T, Savery NJ, Busby SJ, Ross WE, Gaal T, Gourse RL, Thomas MS. 2001. UP element-dependent transcription at the *Escherichia coli rrnB* P1 promoter: positional requirements and role of the RNA polymerase alpha subunit linker. Nucleic Acids Res., 29:4166-78.
- Meyer JE, Schulz GE. 1997. Energy profile of maltooligosaccharide permeation through maltoporin as derived from the structure and from a statistical analysis of saccharide-protein interactions. Protein Sci., 6:1084-91.
- Meyer-Almes FJ, Porschke D. 1997. The cyclic AMP receptor promoter DNA complex: a comparison of crystal and solution structure by quantitative molecular electrooptics. J. Mol. Biol., 269:842-50.
- Miller JH. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
- Mitchell JE, Zheng D, Busby SJ, Minchin SD. 2003. Identification and analysis of 'extended -10' promoters in *Escherichia coli*. Nucleic Acids Res., 31:4689-95.
- Mizuno T. 1997. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. DNA Res., 4:161-8.
- Mizuno T. 1998. His-Asp phosphotransfer signal transduction. J. Biochem. (Tokyo), 123:555-63.
- Mizuno T, Kaneko T, Tabata S. 1996. Compilation of all genes encoding bacterial two-component signal transducers in the genome of the cyanobacterium, *Synechocystis* sp. strain PCC 6803. DNA Res., 3:407-14.
- Moll I, Afonyushkin T, Vytvytska O, Kaberdin VR, Blasi U. 2003a. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. RNA, 9:1308-14.
- Moll I, Leitsch D, Steinhauser T, Blasi U. 2003b. RNA chaperone activity of the Sm-like Hfq protein. EMBO Rep., 4:284-9.
- Moller T, Franch T, Hojrup P, Keene DR, Bachinger HP, Brennan RG, Valentin-Hansen P. 2002. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol. Cell, 9:23-30.
- Mooney RA, Landick R. 2003. Tethering sigma70 to RNA polymerase reveals high *in vivo* activity of sigma factors and σ^{70} -dependent pausing at promoter-distal locations. Genes Dev., 17:2839-51.
- Morita T, El-Kazzaz W, Tanaka Y, Inada T, Aiba H. 2003. Accumulation of glucose 6-phosphate or fructose 6-phosphate is responsible for destabilization of glucose transporter mRNA in *Escherichia coli*. J. Biol. Chem., 278:15608-14.
- Moskovitz J, Poston JM, Berlett BS, Nosworthy NJ, Szczepanowski R, Stadtman ER. 2000. Identification and characterization of a putative active site for peptide methionine sulfoxide reductase (MsrA) and its substrate stereospecificity. J. Biol. Chem., 275:14167-72.
- Muffler A, Barth M, Marschall C, Hengge-Aronis R. 1997a. Heat shock regulation of σ^S turnover: a role for DnaK and relationship between stress responses mediated by σ^S and σ^{32} in *Escherichia coli*. J. Bacteriol., 179:445-52.
- Muffler A, Fischer D, Hengge-Aronis R. 1996. The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli*. Genes Dev., 10:1143-51.
- Mukherjee K, Chatterji D. 1997. Studies on the omega subunit of *Escherichia coli* RNA polymerase--its role in the recovery of denatured enzyme activity. Eur. J. Biochem., 247:884-9.
- Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW, Ebright RH. 2001. Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell, 106:453-63.
- Muller-Hill B. 1998. Some repressors of bacterial transcription. Curr. Opin. Microbiol., 1:145-51.
- Mulligan ME, Hawley DK, Entriken R, McClure WR. 1984. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. Nucleic Acids Res., 12:789-800.
- Murakami KS, Darst SA. 2003. Bacterial RNA polymerases: the whole story. Curr. Opin. Struct. Biol., 13:31-9.
- Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. 2002b. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science, 296:1285-90.
- Murakami KS, Masuda S, Darst SA. 2002a. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science, 296:1280-4.

- Murray HD, Schneider DA, Gourse RL. 2003. Control of rRNA expression by small molecules is dynamic and nonredundant. Mol. Cell, **12**:125-34.
- Murtin C, Engelhorn M, Geiselman J, Boccard F. 1998. A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: re-evaluation of the effective concentration of IHF in the cell. J. Mol. Biol., **284**:949-61.
- Muskhelishvili G, Buckle M, Heumann H, Kahmann R, Travers AA. 1997. FIS activates sequential steps during transcription initiation at a stable RNA promoter. EMBO J., **16**:3655-65.
- Mustaev A, Kashlev M, Lee JY, Polyakov A, Lebedev A, Zalenskaya K, Grachev M, Goldfarb A, Nikiforov V. 1991. Mapping of the priming substrate contacts in the active center of *Escherichia coli* RNA polymerase. J. Biol. Chem., **266**:23927-31.
- Mustaev A, Kozlov M, Markovtsov V, Zaychikov E, Denissova L, Goldfarb A. 1997. Modular organization of the catalytic center of RNA polymerase. Proc. Natl. Acad. Sci. USA, **94**:6641-5.
- N
- Nam TW, Cho SH, Shin D, Kim JH, Jeong JY, Lee JH, Roe JH, Peterkofsky A, Kang SO, Ryu S, Seok YJ. 2001. The *Escherichia coli* glucose transporter enzyme IICB^{Glc} recruits the global repressor Mlc. EMBO J., **20**:491-8.
- Naryshkin N, Revyakin A, Kim Y, Mekler V, Ebright RH. 2000. Structural organization of the RNA polymerase-promoter open complex. Cell, **101**:601-11.
- Nasser W, Schneider R, Travers A, Muskhelishvili G. 2001. CRP modulates fis transcription by alternate formation of activating and repressing nucleoprotein complexes. J. Biol. Chem., **276**:17878-86.
- Negre D, Cortay JC, Galinier A, Sauve P, Cozzone AJ. 1992. Specific interactions between the IclR repressor of the acetate operon of *Escherichia coli* and its operator. J. Mol. Biol., **228**:23-9.
- Nègre D, Oudot C, Prost JF, Murakami K, Ishihama A, Cozzone AJ, Cortay JC. 1998. FruR-mediated transcriptional activation at the *ppsA* promoter of *Escherichia coli*. J. Mol. Biol., **276**:355-65.
- Neiers F, Kriznik A, Boschi-Muller S, Branlant G. 2004. Evidence for a new sub-class of methionine sulfoxide reductases B with an alternative thioredoxin-recognition signature. J. Biol. Chem., **26**:26.
- Nellemann LJ, Holm F, Atlung T, Hansen FG. 1989. Cloning and characterization of the *Escherichia coli* phosphoglycerate kinase (*pgk*) gene. Gene, **77**:185-91.
- Newlands JT, Josaitis CA, Ross W, Gourse RL. 1992. Both fis-dependent and factor-independent upstream activation of the *rrnB* P1 promoter are face of the helix dependent. Nucleic Acids Res., **20**:719-26.
- Nikaïdo H. 1996. Outer membrane. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. Escherichia coli and Salmonella: cellular and molecular biology 2nd edn. Washington DC: American Society for microbiology press. pp 29-47.
- Nikaïdo H, Rosenberg EY. 1983. Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins. J. Bacteriol., **153**:241-52.
- Ninfa EG, Atkinson MR, Kamberov ES, Ninfa AJ. 1993. Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. J. Bacteriol., **175**:7024-32.
- Niu W, Kim Y, Tau G, Heyduk T, Ebright RH. 1996. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. Cell, **87**:1123-34.
- Niu W, Zhou Y, Dong Q, Ebright YW, Ebright RH. 1994. Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP). I. Saturation and alanine-scanning mutagenesis. J. Mol. Biol., **243**:595-602.
- Notley-McRobb L, Ferenci T. 2000. Substrate specificity and signal transduction pathways in the glucose-specific enzyme II (EII^{Glc}) component of the *Escherichia coli* phosphotransferase system. J. Bacteriol., **182**:4437-42.
- Nudler E, Gottesman ME. 2002. Transcription termination and anti-termination in *E. coli*. Genes Cells, **7**:755-68.
- Nuoffer C, Zanolari B, Erni B. 1988. Glucose permease of *Escherichia coli*. The effect of cysteine to serine mutations on the function, stability, and regulation of transport and phosphorylation. J. Biol. Chem., **263**:6647-55.
- O
- Olry A, Boschi-Muller S, Marraud M, Sanglier-Cianferani S, Van Dorsselaar A, Branlant G. 2002. Characterization of the methionine sulfoxide reductase activities of PILB, a probable virulence factor from *Neisseria meningitidis*. J. Biol. Chem., **277**:12016-22.
- Opel ML, Aeling KA, Holmes WM, Johnson RC, Benham CJ, Hatfield GW. 2004. Activation of transcription initiation from a stable RNA promoter by a Fis protein-mediated DNA structural transmission mechanism. Mol. Microbiol., **53**:665-74.
- Oshima T, Aiba H, Masuda Y, Kanaya S, Sugiura M, Wanner BL, Mori H, Mizuno T. 2002.

- Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. Mol. Microbiol., 46:281-91.
- Ovchinnikov Yu A, Monastyrskaya GS, Gubanov VV, Guryev SO, Salomatina IS, Shuvaeva TM, Lipkin VM, Sverdlov ED. 1982. The primary structure of *E. coli* RNA polymerase, Nucleotide sequence of the *rpoC* gene and amino acid sequence of the β' -subunit. Nucleic Acids Res., 10:4035-44.
- Ovchinnikov YA, Lipkin VM, Modyanov NN, Chertov OY, Smirnov YV. 1977. Primary structure of α -subunit of DNA-dependent RNA polymerase from *Escherichia coli*. FEBS Lett., 76:108-11.
- Ovchinnikov YA, Monastyrskaya GS, Gubanov VV, Guryev SO, Chertov O, Modyanov NN, Grinkevich VA, Makarova IA, Marchenko TV, Polovnikova IN, Lipkin VM, Sverdlov ED. 1981. The primary structure of *Escherichia coli* RNA polymerase. Nucleotide sequence of the *rpoB* gene and amino-acid sequence of the β -subunit. Eur. J. Biochem., 116:621-9.
- Owens JT, Miyake R, Murakami K, Chmura AJ, Fujita N, Ishihama A, Meares CF. 1998. Mapping the σ^{70} subunit contact sites on *Escherichia coli* RNA polymerase with a σ^{70} -conjugated chemical protease. Proc. Natl. Acad. Sci. USA, 95:6021-6.
- Ozoline ON, Deev AA, Arkhipova MV. 1997. Non-canonical sequence elements in the promoter structure. Cluster analysis of promoters recognized by *Escherichia coli* RNA polymerase. Nucleic Acids Res., 25:4703-9.
- Ozoline ON, Tsyganov MA. 1995. Structure of open promoter complexes with *Escherichia coli* RNA polymerase as revealed by the DNase I footprinting technique: compilation analysis. Nucleic Acids Res., 23:4533-41.
- P
- Paget MS, Helmann JD. 2003. The σ^{70} family of sigma factors. Genome Biol., 4:3.
- Panaghie G, Aiyar SE, Bobb KL, Hayward RS, de Haseth PL. 2000. Aromatic amino acids in region 2.3 of *Escherichia coli* σ^{70} participate collectively in the formation of an RNA polymerase-promoter open complex. J. Mol. Biol., 299:1217-30.
- Pancholi V, Fischetti VA. 1992. A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity. J. Exp. Med., 176:415-26.
- Park SJ, McCabe J, Turna J, Gunsalus RP. 1994. Regulation of the citrate synthase (*gltA*) gene of *Escherichia coli* in response to anaerobiosis and carbon supply: role of the *arcA* gene product. J. Bacteriol., 176:5086-92.
- Parkinson G, Wilson C, Gunasekera A, Ebricht YW, Ebricht RE, Berman HM. 1996. Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. J. Mol. Biol., 260:395-408.
- Passner JM, Schultz SC, Steitz TA. 2000. Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution. J. Mol. Biol., 304:847-59.
- Passner JM, Steitz TA. 1997. The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. Proc. Natl. Acad. Sci. USA, 94:2843-7.
- Pautsch A, Schulz GE. 1998. Structure of the outer membrane protein A transmembrane domain. Nat. Struct. Biol., 5:1013-7.
- Pemberton IK, Muskhelishvili G, Travers AA, Buckle M. 2000. The G+C-rich discriminator region of the *tyrT* promoter antagonises the formation of stable preinitiation complexes. J. Mol. Biol., 299:859-64.
- Pemberton IK, Muskhelishvili G, Travers AA, Buckle M. 2002. FIS modulates the kinetics of successive interactions of RNA polymerase with the core and upstream regions of the *tyrT* promoter. J. Mol. Biol., 318:651-63.
- Perez-Martin J, de Lorenzo V. 1997. Clues and consequences of DNA bending in transcription. Ann. Rev. Microbiol., 51:593-628.
- Perez-Martin J, Espinosa M. 1994. Correlation between DNA bending and transcriptional activation at a plasmid promoter. J. Mol. Biol., 241:7-17.
- Perez-Martin J, Rojo F, de Lorenzo V. 1994. Promoter responsive to DNA bending: a common theme in procaryotic gene expression. Microbiol. Rev., 58:268-290.
- Perez-Rueda E, Collado-Vides J. 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res., 28:1838-47.
- Petersen C, Moller LB. 2000. Invariance of the nucleoside triphosphate pools of *Escherichia coli* with growth rate. J. Biol. Chem., 275:3931-5.
- Pettijohn DE. 1996. The nucleoid. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. Escherichia coli and Salmonella: cellular and molecular biology 2nd edn. Washington DC: American Society for microbiology press. pp 158-166.
- Plumbridge J. 1996. How to achieve constitutive expression of a gene within an inducible operon: the example of the *nagC* gene of *Escherichia coli*. J. Bacteriol., 178:2629-36.
- Plumbridge J. 1998a. Control of the expression of the *manXYZ* operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. Mol. Microbiol., 27:369-80.

- Plumbridge J. 1998b. Expression of *ptsG*, the gene for the major glucose PTS transporter in *Escherichia coli*, is repressed by Mlc and induced by growth on glucose. Mol. Microbiol., 29:1053-63.
- Plumbridge J. 1999. Expression of the phosphotransferase system both mediates and is mediated by *mlc* regulation in *Escherichia coli*. Mol. Microbiol., 33:260-73.
- Plumbridge J. 2001. DNA binding sites for the Mlc and NagC proteins: regulation of *nagE*, encoding the N-acetylglucosamine-specific transporter in *Escherichia coli*. Nucleic Acids Res., 29:506-14.
- Plumbridge J. 2001. Regulation of PTS gene expression by the homologous transcriptional regulators, Mlc and NagC, in *Escherichia coli* (or how two similar repressors can behave differently). J. Mol. Microbiol. Biotechnol., 3:371-80.
- Plumbridge J. 2002. Regulation of gene expression in the PTS in *Escherichia coli*: the role and interactions of Mlc. Curr. Opin. Microbiol., 5:187-93.
- Plumbridge J, Kolb A. 1991. CAP and Nag repressor binding to the regulatory regions of the *nagE-B* and *manX* genes of *Escherichia coli*. J. Mol. Biol., 217:661-79.
- Plumbridge J, Kolb A. 1993. DNA loop formation between Nag repressor molecules bound to its two operator sites is necessary for repression of the *nag* regulon of *Escherichia coli* *in vivo*. Mol. Microbiol., 10:973-81.
- Plumbridge J, Kolb A. 1998. DNA bending and expression of the divergent *nagE-B* operons. Nucleic Acids Res., 26:1254-60.
- Plumbridge J, Soll D. 1989. Characterization of cis-acting mutations which increase expression of a *glnS-lacZ* fusion in *Escherichia coli*. Mol. Gen. Genet., 216:113-9.
- Plumbridge JA. 1990. Induction of the *nag* regulon of *Escherichia coli* by N-acetylglucosamine and glucosamine: role of the cyclic AMP-catabolite activator protein complex in expression of the regulon. J. Bacteriol., 172:2728-35.
- Plumbridge JA. 1991. Repression and induction of the *nag* regulon of *Escherichia coli* K-12: the roles of *nagC* and *nagA* in maintenance of the uninduced state. Mol. Microbiol., 5:2053-62.
- Polyakov A, Severinova E, Darst SA. 1995. Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. Cell, 83:365-73.
- Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev., 57:543-94.
- Postma PW, Lengeler JW, Jacobson GR. 1996. Phosphoenolpyruvate:carbohydrate phosphotransferase systems. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. *Escherichia coli* and *Salmonella*: cellular and molecular biology 2nd edn. Washington DC: American Society for microbiology press. pp 1149-1174.
- Powell BS, Rivas MP, Court DL, Nakamura Y, Turnbough CL, Jr. 1994. Rapid confirmation of single copy lambda prophage integration by PCR. Nucl. Acids Res., 22:5765-6.
- Pratt LA, Hsing W, Gibson KE, Silhavy TJ. 1996. From acids to *osmZ*: multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli*. Mol. Microbiol., 20:911-7.
- Prost JF, Nègre D, Oudot C, Murakami K, Ishihama A, Cozzzone AJ, Cortay JC. 1999. Cra-dependent transcriptional activation of the *icd* gene of *Escherichia coli*. J. Bacteriol., 181:893-8.
- Pruss BM, Liu X, Hendrickson W, Matsumura P. 2001. FlhD/FlhC-regulated promoters analyzed by gene array and *lacZ* gene fusions. FEMS Microbiol. Lett., 197:91-7.
- Pruss GJ, Drlica K. 1989. DNA supercoiling and prokaryotic transcription. Cell, 56:521-3.
- Pyles EA, Chin AJ, Lee JC. 1998a. *Escherichia coli* cAMP receptor protein-DNA complexes. 1. Energetic contributions of half-sites and flanking sequences in DNA recognition. Biochemistry, 37:5194-200.
- Pyles EA, Lee JC. 1996. Mode of selectivity in cyclic AMP receptor protein-dependent promoters in *Escherichia coli*. Biochemistry, 35:1162-72.
- Pyles EA, Lee JC. 1998b. *Escherichia coli* cAMP receptor protein-DNA complexes. 2. Structural asymmetry of DNA bending. Biochemistry, 37:5201-10.

R

- Raleigh EA, Murray NE, Revel H, Blumenthal RM, Westaway D, Reith AD, Rigby PW, Elhai J, Hanahan D. 1988. McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. Nucleic Acids Res., 16:1563-75.
- Ramseier TM, Bledig S, Michotey V, Feghali R, Saier MH, Jr. 1995. The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. Mol. Microbiol., 16:1157-69.
- Ramseier TM, Nègre D, Cortay JC, Scarabel M, Cozzzone AJ, Saier MH, Jr. 1993. *In vitro* binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru*, *pps*, *ace*, *pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*. J. Mol. Biol., 234:28-44.
- Rao L, Ross W, Appleman JA, Gaal T, Leirimo S, Schlax PJ, Record MT, Jr., Gourse RL. 1994. Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J. Mol. Biol., 235:1421-35.

- Record MT, Reznikoff WS, Craig ML, McQuade KL, Schlax PJ. 1996. *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. *Escherichia coli and Salmonella: cellular and molecular biology 2nd edn*. Washington DC: American Society for microbiology press. pp 792-820.
- Reddy P, Kamireddi M. 1998. Modulation of *Escherichia coli* adenylyl cyclase activity by catalytic-site mutants of protein IIA^{Glc} of the phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.*, 180:732-6.
- Reitzer L. 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu. Rev. Microbiol.*, 57:155-76.
- Reitzer L, Schneider BL. 2001. Metabolic context and possible physiological themes of σ^{54} -dependent genes in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.*, 65:422-44.
- Revyakin A, Ebricht RH, Strick TR. 2004. Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. *Proc. Natl. Acad. Sci. USA*, 101:4776-80.
- Rhodium VA, Busby SJ. 2000. Transcription activation by the *Escherichia coli* cyclic AMP receptor protein: determinants within activating region 3. *J. Mol. Biol.*, 299:295-310.
- Rhodium VA, West DM, Webster CL, Busby SJ, Savery NJ. 1997. Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res.*, 25:326-32.
- Richardson JP, Greenblatt J. 1996. Control of RNA chain elongation and termination. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. *Escherichia coli and Salmonella: cellular and molecular biology 2nd edn*. Washington DC: American Society for microbiology press. pp 882-843.
- Richet E. 1996. On the role of the multiple regulatory elements involved in the activation of the *Escherichia coli malEp* promoter. *J. Mol. Biol.*, 264:852-62.
- Richet E. 2000. Synergistic transcription activation: a dual role for CRP in the activation of an *Escherichia coli* promoter depending on MalT and CRP. *EMBO J.*, 19:5222-32.
- Ring BZ, Yarnell WS, Roberts JW. 1996. Function of *E. coli* RNA polymerase sigma factor σ^{70} in promoter-proximal pausing. *Cell*, 86:485-93.
- Rippe K, Mucke N, Schulz A. 1998. Association states of the transcription activator protein NtrC from *E. coli* determined by analytical ultracentrifugation. *J. Mol. Biol.*, 278:915-33.
- Rivetti C, Guthold M, Bustamante C. 1999. Wrapping of DNA around the *E. coli* RNA polymerase open promoter complex. *EMBO J.*, 18:4464-75.
- Robas N, Branlant C. 1994. The expression of the penicillin G amidase gene of *Escherichia coli* by primer extension analysis. *Curr. Microbiol.*, 29:263-8.
- Roberts DL, Bennett DW, Forst SA. 1994. Identification of the site of phosphorylation on the osmosensor, EnvZ, of *Escherichia coli*. *J. Biol. Chem.*, 269:8728-33.
- Robillard GT, Broos J. 1999. Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim. Biophys. Acta*, 1422:73-104.
- Robinson VL, Buckler DR, Stock AM. 2000. A tale of two components: a novel kinase and a regulatory switch. *Nat. Struct. Biol.*, 7:626-33.
- Rojo F. 1999. Repression of transcription initiation in bacteria. *J. Bacteriol.*, 181:2987-91.
- Rojo F. 2001. Mechanisms of transcriptional repression. *Curr. Opin. Microbiol.*, 4:145-51.
- Rombel I, North A, Hwang I, Wyman C, Kustu S. 1998. The bacterial enhancer-binding protein NtrC as a molecular machine. *Cold Spring Harb. Symp. Quant. Biol.*, 63:157-66.
- Romeo T. 1998. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.*, 29:1321-30.
- Ross W, Aiyar SE, Salomon J, Gourse RL. 1998. *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *J. Bacteriol.*, 180:5375-83.
- Ross W, Ernst A, Gourse RL. 2001. Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev.*, 15:491-506.
- Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, Severinov K, Gourse RL. 1993. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science*, 262:1407-13.
- Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL. 2003. An intersubunit contact stimulating transcription initiation by *E. coli* RNA polymerase: interaction of the α C-terminal domain and σ region 4. *Genes Dev.*, 17:1293-307.
- Ross W, Thompson JF, Newlands JT, Gourse RL. 1990. *E. coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J.*, 9:3733-42.
- Rouvière-Yaniv J, Kiseleva E, Bensaid A, Almeida A, Drlica K. 1992. Protein HU and bacterial DNA supercoiling. In: Mohan S, Dow C, Cole A, eds. *Prokaryotic Structure and Function: A New Perspective*. Cambridge: Cambridge University Press. pp 17-39.
- Roy A, Danchin A. 1982. The *cya* locus of *Escherichia coli* K12: organization and gene products. *Mol. Gen. Genet.*, 188:465-71.

- Roy S, Garges S, Adhya S. 1998. Activation and repression of transcription by differential contact: two sides of a coin. *J. Biol. Chem.*, 273:14059-62.
- Roy S, Lim HM, Liu M, Adhya S. 2004. Asynchronous basepair openings in transcription initiation: CRP enhances the rate-limiting step. *EMBO J.*, 23:869-75.
- Ruijter GJ, van Meurs G, Verwey MA, Postma PW, van Dam K. 1992. Analysis of mutations that uncouple transport from phosphorylation in enzyme II^{Glc} of the *Escherichia coli* phosphoenolpyruvate-dependent phosphotransferase system. *J. Bacteriol.*, 174:2843-50.
- Ryu S. 1995. Gene expression and mRNA. *Nature*, 376:560.
- Ryu S, Garges S. 1994. Promoter switch in the *Escherichia coli pts* operon. *J. Biol. Chem.*, 269:4767-72.
- S
- Saecker RM, Tsodikov OV, McQuade KL, Schlax PE, Jr., Capp MW, Record MT, Jr. 2002. Kinetic studies and structural models of the association of *E. coli* σ^{70} RNA polymerase with the lambdaP(R) promoter: large scale conformational changes in forming the kinetically significant intermediates. *J. Mol. Biol.*, 319:649-71.
- Saier MH, Jr., Novotny MJ, Comeau-Fuhrman D, Osumi T, Desai JD. 1983. Cooperative binding of the sugar substrates and allosteric regulatory protein (enzyme III^{Glc} of the phosphotransferase system) to the lactose and melibiose permeases in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.*, 155:1351-7.
- Saier MH, Jr., Ramseier TM. 1996. The catabolite repressor/activator (Cra) protein of enteric bacteria. *J. Bacteriol.*, 178:3411-7.
- Saier MH, Jr., Straud H, Massman LS, Judice JJ, Newman MJ, Feucht BU. 1978. Permease-specific mutations in *Salmonella typhimurium* and *Escherichia coli* that release the glycerol, maltose, melibiose, and lactose transport systems from regulation by the phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.*, 133:1358-67.
- Saier MH, Ramseier TM, Reizer J. 1996. Regulation of carbon utilization. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, eds. *Escherichia coli and Salmonella: cellular and molecular biology 2nd edn*. Washington DC: American Society for microbiology press. pp 1325-1343.
- Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C, Collado-Vides J. 2004. RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res.*, 32 Database issue:D303-6.
- Sambrook J, Fritsch FF, Maniatis T. 1989. *Molecular cloning: a laboratory manual, 2nd ed*. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
- Sanderson A, Mitchell J, Minchin S, Busby S. 2003. Substitutions in the *Escherichia coli* RNA polymerase σ^{70} factor that affect recognition of extended -10 elements at promoters. *FEBS Lett.*, 544:199-205.
- Sasse-Dwight S, Gralla JD. 1990. Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor σ^{54} . *Cell*, 62:945-54.
- Savery NJ, Lloyd GS, Busby SJ, Thomas MS, Ebright RH, Gourse RL. 2002. Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase α subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters. *J. Bacteriol.*, 184:2273-80.
- Sawers G. 1999. The aerobic/anaerobic interface. *Curr. Opin. Microbiol.*, 2:181-7.
- Schickor P, Metzger W, Werel W, Lederer H, Heumann H. 1990. Topography of intermediates in transcription initiation of *E. coli*. *EMBO J.*, 9:2215-20.
- Schirmer T. 1998. General and specific porins from bacterial outer membranes. *J. Struct. Biol.*, 121:101-9.
- Schirmer T, Keller TA, Wang YF, Rosenbusch JP. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science*, 267:512-4.
- Schlapfer BS, Zuber H. 1992. Cloning and sequencing of the genes encoding glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase (*gap* operon) from mesophilic *Bacillus megaterium*: comparison with corresponding sequences from thermophilic *Bacillus stearothermophilus*. *Gene*, 122:53-62.
- Schlax PJ, Capp MW, Record MT, Jr. 1995. Inhibition of transcription initiation by *lac* repressor. *J. Mol. Biol.*, 245:331-50.
- Schneider DA, Gaal T, Gourse RL. 2002. NTP-sensing by rRNA promoters in *Escherichia coli* is direct. *Proc. Natl. Acad. Sci. USA*, 99:8602-7.
- Schneider DA, Gourse RL. 2004. Relationship between growth rate and ATP concentration in *Escherichia coli*: a bioassay for available cellular ATP. *J. Biol. Chem.*, 279:8262-8.
- Schneider DA, Ross W, Gourse RL. 2003. Control of rRNA expression in *Escherichia coli*. *Curr. Opin. Microbiol.*, 6:151-6.
- Schneider R, Travers A, Kutateladze T, Muskhelishvili G. 1999. A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.*, 34:953-64.

- Schroder I, Darie S, Gunsalus RP. 1993. Activation of the *Escherichia coli* nitrate reductase (*narGHJ*) operon by NarL and Fnr requires integration host factor. *J. Biol. Chem.*, 268:771-4.
- Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science*, 253:1001-7.
- Schulz GE. 1996. Porins: general to specific, native to engineered passive pores. *Curr. Opin. Struct. Biol.*, 6:485-90.
- Sclavi B, Woodson S, Sullivan M, Chance MR, Brenowitz M. 1997. Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein-DNA interactions and RNA folding. *J. Mol. Biol.*, 266:144-59.
- Seitz S, Lee SJ, Penner C, Boos W, Plumbridge J. 2003. Analysis of the interaction between the global regulator Mlc and EIIB^{Glc} of the glucose-specific phosphotransferase system in *Escherichia coli*. *J. Biol. Chem.*, 278:10744-51.
- Seok YJ, Sun J, Kaback HR, Peterkofsky A. 1997b. Topology of allosteric regulation of lactose permease. *Proc. Natl. Acad. Sci. USA*, 94:13515-9.
- Seok YJ, Zhu PP, Koo BM, Peterkofsky A. 1998. Autophosphorylation of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system requires dimerization. *Biochem. Biophys. Res. Commun.*, 250:381-4.
- Severinov K, Fenyo D, Severinova E, Mustaev A, Chait BT, Goldfarb A, Darst SA. 1994. The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *J. Biol. Chem.*, 269:20826-8.
- Sharov VS, Ferrington DA, Squier TC, Schoneich C. 1999. Diastereoselective reduction of protein-bound methionine sulfoxide by methionine sulfoxide reductase. *FEBS Lett.*, 455:247-50.
- Sharp MM, Chan CL, Lu CZ, Marr MT, Nechaev S, Merritt EW, Severinov K, Roberts JW, Gross CA. 1999. The interface of σ with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev.*, 13:3015-26.
- Shin D, Cho N, Heu S, Ryu S. 2003. Selective regulation of *ptsG* expression by Fis. Formation of either activating or repressing nucleoprotein complex in response to glucose. *J. Biol. Chem.*, 278:14776-81.
- Shin M, Kang S, Hyun SJ, Fujita N, Ishihama A, Valentin-Hansen P, Choy HE. 2001. Repression of *deoP2* in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *EMBO J.*, 20:5392-9.
- Shin S, Park C. 1995. Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. *J. Bacteriol.*, 177:4696-702.
- Siebenlist U, Simpson RB, Gilbert W. 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell*, 20:269-81.
- Siebold C, Flukiger K, Beutler R, Erni B. 2001. Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). *FEBS Lett.*, 504:104-11.
- Siegele DA, Hu JC, Walter WA, Gross CA. 1989. Altered promoter recognition by mutant forms of the σ^{70} subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.*, 206:591-603.
- Silversmith RE, Bourret RB. 1999. Throwing the switch in bacterial chemotaxis. *Trends Microbiol.*, 7:16-22.
- Simms SA, Keane MG, Stock J. 1985. Multiple forms of the CheB methyltransferase in bacterial chemosensing. *J. Biol. Chem.*, 260:10161-8.
- Simons RW, Houtman F, Kleckner N. 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*, 53:85-96.
- Sirover MA. 1996. Emerging new functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells. *Life Sci.*, 58:2271-7.
- Sledjeski DD, Gupta A, Gottesman S. 1996. The small RNA, *DsrA*, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. *EMBO J.*, 15:3993-4000.
- Sondej M, Sun J, Seok YJ, Kaback HR, Peterkofsky A. 1999. Deduction of consensus binding sequences on proteins that bind IIA^{Glc} of the phosphoenolpyruvate:sugar phosphotransferase system by cysteine scanning mutagenesis of *Escherichia coli* lactose permease. *Proc. Natl. Acad. Sci. USA*, 96:3525-30.
- Sorenson MK, Ray SS, Darst SA. 2004. Crystal structure of the flagellar sigma/anti-sigma complex σ^{28} /FlgM reveals an intact sigma factor in an inactive conformation. *Mol. Cell*, 14:127-38.
- Soutourina O, Kolb A, Krin E, Laurent-Winter C, Rimsky S, Danchin A, Bertin P. 1999. Multiple control of flagellum biosynthesis in *Escherichia coli*: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the *flhDC* master operon. *J. Bacteriol.*, 181:7500-8.
- Soutourina OA, Bertin PN. 2003. Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiol. Rev.*, 27:505-23.
- Steck TR, Franco RJ, Wang JY, Drlica K. 1993. Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins. *Mol. Microbiol.*, 10:473-81.
- Stock AM, Robinson VL, Goudreau PN. 2000. Two-component signal transduction. *Annu. Rev. Biochem.*, 69:183-215.
- Stock JB, Surette MG. 1996. Chemotaxis. In: Neidhart F, Curtiss R, Ingraham JL, Lin ECC,

T

- Low KB, Magasanik B, eds. *Escherichia coli and Salmonella: cellular and molecular biology 2nd edn*. Washington DC: American Society for microbiology press. pp 1103-1129.
- Storz G, Opdyke JA, Zhang A. 2004. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.*, **7**:140-4.
- Strainic MG, Jr., Sullivan JJ, Velevis A, deHaseth PL. 1998. Promoter recognition by *Escherichia coli* RNA polymerase: effects of the UP element on open complex formation and promoter clearance. *Biochemistry*, **37**:18074-80.
- Straney DC, Crothers DM. 1987. A stressed intermediate in the formation of stably initiated RNA chains at the *Escherichia coli lac UV5* promoter. *J. Mol. Biol.*, **193**:267-78.
- Studemann A, Noirclerc-Savoye M, Klauck E, Becker G, Schneider D, Hengge R. 2003. Sequential recognition of two distinct sites in σ^S by the proteolytic targeting factor RssB and ClpX. *EMBO J.*, **22**:4111-20.
- Studholme DJ, Buck M. 2000. The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol. Lett.*, **186**:1-9.
- Studitsky V, Brodolin K, Liu Y, Mirzabekov A. 2001. Topography of *lacUV5* initiation complexes. *Nucleic Acids Res.*, **29**:854-61.
- Stulke J, Hillen W. 1999. Carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.*, **2**:195-201.
- Su W, Porter S, Kustu S, Echols H. 1990. DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc. Natl. Acad. Sci. USA*, **87**:5504-8.
- Suzuki T, Ueguchi C, Mizuno T. 1996. H-NS regulates OmpF expression through *micF* antisense RNA in *Escherichia coli*. *J. Bacteriol.*, **178**:3650-3.
- Swanson RV, Bourret RB, Simon MI. 1993. Intermolecular complementation of the kinase activity of CheA. *Mol. Microbiol.*, **8**:435-41.
- Sweetser D, Nonet M, Young RA. 1987. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Natl. Acad. Sci. USA*, **84**:1192-6.
- Syed A, Gralla JD. 1998. Identification of an N-terminal region of σ^{54} required for enhancer responsiveness. *J. Bacteriol.*, **180**:5619-25.
- Szoke PA, Allen TL, deHaseth PL. 1987. Promoter recognition by *Escherichia coli* RNA polymerase: effects of base substitutions in the -10 and -35 regions. *Biochemistry*, **26**:6188-94.
- Szurmant H, Ordal GW. 2004. Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.*, **68**:301-19.
- Tabor S, Richardson CC. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA*, **84**:4767-71.
- Tagami H, Aiba H. 1999. An inactive open complex mediated by an UP element at *Escherichia coli* promoters. *Proc. Natl. Acad. Sci. USA*, **96**:7202-7.
- Takahashi H, Inada T, Postma P, Aiba H. 1998. CRP down-regulates adenylate cyclase activity by reducing the level of phosphorylated IIA^{Glc}, the glucose-specific phosphotransferase protein, in *Escherichia coli*. *Mol. Gen. Genet.*, **259**:317-26.
- Takeda S, Fujisawa Y, Matsubara M, Aiba H, Mizuno T. 2001. A novel feature of the multistep phosphorelay in *Escherichia coli*: a revised model of the RcsC \rightarrow YojN \rightarrow RcsB signalling pathway implicated in capsular synthesis and swarming behaviour. *Mol. Microbiol.*, **40**:440-50.
- Takeda S, Matsushika A, Mizuno T. 1999. Repression of the gene encoding succinate dehydrogenase in response to glucose is mediated by the EIICB^{Glc} protein in *Escherichia coli*. *J. Biochem. (Tokyo)*, **126**:354-60.
- Tan GS, Kelly P, Kim J, Wartell RM. 1991. Comparison of cAMP receptor protein (CRP) and a cAMP-independent form of CRP by Raman spectroscopy and DNA binding. *Biochemistry*, **30**:5076-80.
- Tanaka Y, Itoh F, Kimata K, Aiba H. 2004. Membrane localization itself but not binding to IICB is directly responsible for the inactivation of the global repressor Mlc in *Escherichia coli*. *Mol. Microbiol.*, **53**:941-51.
- Tanaka Y, Kimata K, Aiba H. 2000. A novel regulatory role of glucose transporter of *Escherichia coli*: membrane sequestration of a global repressor Mlc. *EMBO J.*, **19**:5344-52.
- Tanaka Y, Kimata K, Inada T, Tagami H, Aiba H. 1999. Negative regulation of the *pts* operon by Mlc: mechanism underlying glucose induction in *Escherichia coli*. *Genes Cells*, **4**:391-9.
- Tang H, Severinov K, Goldfarb A, Fenyo D, Chait B, Ebright RH. 1994. Location, structure, and function of the target of a transcriptional activator protein. *Genes Dev.*, **8**:3058-67.
- Tardat B, Touati D. 1993. Iron and oxygen regulation of *Escherichia coli* MnSOD expression: competition between the global regulators Fur and ArcA for binding to DNA. *Mol. Microbiol.*, **9**:53-63.
- Tavormina PL, Landick R, Gross CA. 1996. Isolation, purification, and *in vitro* characterization of recessive-lethal-mutant RNA polymerases from *Escherichia coli*. *J. Bacteriol.*, **178**:5263-71.
- Taylor M, Butler R, Chambers S, Casimiro M, Badii F, Merrick M. 1996. The RpoN-box motif of the

- RNA polymerase sigma factor σ^N plays a role in promoter recognition. Mol. Microbiol., 22:1045-54.
- Temiakov D, Patlan V, Anikin M, McAllister WT, Yokoyama S, Vassilyev DG. 2004. Structural basis for substrate selection by T7 RNA polymerase. Cell, 116:381-91.
- Tete-Favier F, Cobessi D, Boschi-Muller S, Azza S, Branlant G, Aubry A. 2000. Crystal structure of the *Escherichia coli* peptide methionine sulphoxide reductase at 1.9 Å resolution. Structure Fold. Des., 8:1167-78.
- Tete-Favier F, Cobessi D, Leonard GA, Azza S, Talfournier F, Boschi-Muller S, Branlant G, Aubry A. 2000. Crystallization and preliminary X-ray diffraction studies of the peptide methionine sulfoxide reductase from *Escherichia coli*. Acta Crystallogr. D Biol. Crystallogr., 56:1194-7.
- Thomson J, Gerstenberger PD, Goldberg DE, Gociar E, Orozco de Silva A, Fraenkel DG. 1979. ColE1 hybrid plasmids for *Escherichia coli* genes of glycolysis and the hexose monophosphate shunt. J. Bacteriol., 137:502-6.
- Titgemeyer F, Mason RE, Saier MH, Jr. 1994. Regulation of the raffinose permease of *Escherichia coli* by the glucose-specific enzyme IIA of the phosphoenolpyruvate:sugar phosphotransferase system. J. Bacteriol., 176:543-6.
- Tomoyasu T, Gamer J, Bukau B, Kanemori M, Mori H, Rutman AJ, Oppenheim AB, Yura T, Yamanaka K, Niki H, et al. 1995. *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor σ^{32} . EMBO J., 14:2551-60.
- Toulkhonov I, Landick R. 2003. The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. Mol. Cell, 12:1125-36.
- Travers A, Muskhelishvili G. 1998. DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. J. Mol. Biol., 279:1027-43.
- Travers AA. 1980. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. J. Bacteriol., 141:973-6.
- Tsujikawa L, Strainic MG, Watrob H, Barkley MD, DeHaseth PL. 2002. RNA polymerase alters the mobility of an A-residue crucial to polymerase-induced melting of promoter DNA. Biochemistry, 41:15334-41.
- V
- Valentin-Hansen P, Sogaard-Andersen L, Pedersen H. 1996. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. Mol. Microbiol., 20:461-6.
- Vassilieva IM, Nikulin AD, Blasi U, Moll I, Garber MB. 2003. Crystallization of Hfq protein: a bacterial gene-expression regulator. Acta Crystallogr. D Biol. Crystallogr., 59:1061-3.
- Vassilyev DG, Sekine S, Laptenko O, Lee J, Vassilyeva MN, Borukhov S, Yokoyama S. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. Nature, 417:712-9.
- Vo NV, Hsu LM, Kane CM, Chamberlin MJ. 2003a. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 2. Formation and characterization of two distinct classes of initial transcribing complexes. Biochemistry, 42:3787-97.
- Vo NV, Hsu LM, Kane CM, Chamberlin MJ. 2003b. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. Biochemistry, 42:3798-811.
- Vogt J, Schulz GE. 1999. The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. Structure Fold. Des., 7:1301-9.
- Voskuil MI, Chambliss GH. 2002. The TRTGn motif stabilizes the transcription initiation open complex. J. Mol. Biol., 322:521-32.
- Voskuil MI, Voepel K, Chambliss GH. 1995. The -16 region, a vital sequence for the utilization of a promoter in *Bacillus subtilis* and *Escherichia coli*. Mol. Microbiol., 17:271-9.
- Vuthoori S, Bowers CW, McCracken A, Dombroski AJ, Hinton DM. 2001. Domain 1.1 of the σ^{70} subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. J. Mol. Biol., 309:561-72.
- W
- Waldburger C, Gardella T, Wong R, Susskind MM. 1990. Changes in conserved region 2 of *Escherichia coli* σ^{70} affecting promoter recognition. J. Mol. Biol., 215:267-76.
- Walker KA, Atkins CL, Osuna R. 1999. Functional determinants of the *Escherichia coli* *fis* promoter: roles of -35, -10, and transcription initiation regions in the response to stringent control and growth phase-dependent regulation. J. Bacteriol., 181:1269-80.
- Walsh NP, Alba BM, Bose B, Gross CA, Sauer RT. 2003. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. Cell, 113:61-71.
- Wang G, Louis JM, Sondej M, Seok YJ, Peterkofsky A, Clore GM. 2000. Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA^{Glucose} of the

- Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system. EMBO J., 19:5635-49.
- Wang JY, Syvanen M. 1992. DNA twist as a transcriptional sensor for environmental changes. Mol. Microbiol., 6:1861-6.
- Wang Y, Severinov K, Loizos N, Fenyo D, Heyduk E, Heyduk T, Chait BT, Darst SA. 1997a. Determinants for *Escherichia coli* RNA polymerase assembly within the β subunit. J. Mol. Biol., 270:648-62.
- Warne SE, deHaseth PL. 1993. Promoter recognition by *Escherichia coli* RNA polymerase. Effects of single base pair deletions and insertions in the spacer DNA separating the -10 and -35 regions are dependent on spacer DNA sequence. Biochemistry, 32:6134-40.
- Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. 2001. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev., 15:1637-51.
- Wedel A, Kustu S. 1995. The bacterial enhancer-binding protein NTRC is a molecular machine: ATP hydrolysis is coupled to transcriptional activation. Genes Dev., 9:2042-52.
- Wei BL, Brun-Zinkernagel AM, Simecka JW, Pruss BM, Babitzke P, Romeo T. 2001. Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of *Escherichia coli*. Mol. Microbiol., 40:245-56.
- Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T. 2003. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. Mol. Microbiol., 48:657-70.
- Weiss DS, Batut J, Klose KE, Keener J, Kustu S. 1991. The phosphorylated form of the enhancer-binding protein NTRC has an ATPase activity that is essential for activation of transcription. Cell, 67:155-67.
- Weissbach H, Etienne F, Hoshi T, Heinemann SH, Lowther WT, Matthews B, St John G, Nathan C, Brot N. 2002. Peptide methionine sulfoxide reductase: structure, mechanism of action, and biological function. Arch. Biochem. Biophys., 397:172-8.
- Welch M, Oosawa K, Aizawa S, Eisenbach M. 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. Proc. Natl. Acad. Sci. USA, 90:8787-91.
- Welz D, Braun V. 1998. Ferric citrate transport of *Escherichia coli*: functional regions of the FecR transmembrane regulatory protein. J. Bacteriol., 180:2387-94.
- West AH, Stock AM. 2001. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem. Sci., 26:369-76.
- Wilson C, Dombroski AJ. 1997. Region 1 of σ^{70} is required for efficient isomerization and initiation of transcription by *Escherichia coli* RNA polymerase. J. Mol. Biol., 267:60-74.
- Wilson TH, Ding PZ. 2001. Sodium-substrate cotransport in bacteria. Biochim. Biophys. Acta, 1505:121-30.
- Wilson TH, Yunker PL, Hansen CL. 1990. Lactose transport mutants of *Escherichia coli* resistant to inhibition by the phosphotransferase system. Biochim. Biophys. Acta, 1029:113-6.
- Wolanin PM, Thomason PA, Stock JB. 2002. Histidine protein kinases: key signal transducers outside the animal kingdom. Genome Biol., 3:REVIEWS3013.
- Won HS, Yamazaki T, Lee TW, Jee JG, Yoon MK, Park SH, Otomo T, Aiba H, Kyogoku Y, Lee BJ. 2000. Backbone NMR assignments of a high molecular weight protein (47 kDa), cyclic AMP receptor protein (apo-CRP). J. Biomol. NMR, 16:79-80.
- Wong C, Gralla JD. 1992. A role for the acidic trimer repeat region of transcription factor σ^{54} in setting the rate and temperature dependence of promoter melting *in vivo*. J. Biol. Chem., 267:24762-8.
- Wood JM. 1999. Osmosensing by bacteria: signals and membrane-based sensors. Microbiol. Mol. Biol. Rev., 63:230-62.
- Wosten MM. 1998. Eubacterial sigma-factors. FEMS Microbiol. Rev., 22:127-50.
- Wu HM, Crothers DM. 1984. The locus of sequence-directed and protein-induced DNA bending. Nature, 308:509-13.
- Y
- Yamamoto K, Ishihama A. 2003. Two different modes of transcription repression of the *Escherichia coli* acetate operon by IclR. Mol. Microbiol., 47:183-94.
- Yamamoto K, Nagura R, Tanabe H, Fujita N, Ishihama A, Utsumi R. 2000. Negative regulation of the *bolA1p* of *Escherichia coli* K-12 by the transcription factor OmpR for osmolarity response genes. FEMS Microbiol. Lett., 186:257-62.
- Yamamoto K, Ogasawara H, Fujita N, Utsumi R, Ishihama A. 2002. Novel mode of transcription regulation of divergently overlapping promoters by PhoP, the regulator of two-component system sensing external magnesium availability. Mol. Microbiol., 45:423-38.
- Yamashino T, Ueguchi C, Mizuno T. 1995. Quantitative control of the stationary phase-specific sigma factor, σ^S , in *Escherichia coli*: involvement of the nucleoid protein H-NS. EMBO J., 14:594-602.
- Yang Y, Zhao G, Man TK, Winkler ME. 1998. Involvement of the *gapA*- and *epd* (*gapB*)-encoded dehydrogenases in pyridoxal 5'-

- phosphate coenzyme biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.*, 180:4294-9.
- Yanisch-Perron C, Vieira J, Messing J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33:103-19.
- Yin YW, Steitz TA. 2004. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, 116:393-404.
- Yoshida T, Cai S, Inouye M. 2002. Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator. *Mol. Microbiol.*, 46:1283-94.
- Young BA, Gruber TM, Gross CA. 2002. Views of transcription initiation. *Cell*, 109:417-20.
- Yun M, Park CG, Kim JY, Park HW. 2000. Structural analysis of glyceraldehyde 3-phosphate dehydrogenase from *Escherichia coli*: direct evidence of substrate binding and cofactor-induced conformational changes. *Biochemistry*, 39:10702-10.
- Zhou Y, Busby S, Ebright RH. 1993b. Identification of the functional subunit of a dimeric transcription activator protein by use of oriented heterodimers. *Cell*, 73:375-9.
- Zhou Y, Merkel TJ, Ebright RH. 1994a. Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP). II. Role at Class I and class II CAP-dependent promoters. *J. Mol. Biol.*, 243:603-10.
- Zhou Y, Pendergrast PS, Bell A, Williams R, Busby S, Ebright RH. 1994b. The functional subunit of a dimeric transcription activator protein depends on promoter architecture. *EMBO J.*, 13:4549-57.
- Zhou Y, Zhang X, Ebright RH. 1993a. Identification of the activating region of catabolite gene activator protein (CAP): isolation and characterization of mutants of CAP specifically defective in transcription activation. *Proc. Natl. Acad. Sci. USA*, 90:6081-5.

Z

- Zaychikov E, Martin E, Denissova L, Kozlov M, Markovtsov V, Kashlev M, Heumann H, Nikiforov V, Goldfarb A, Mustaev A. 1996. Mapping of catalytic residues in the RNA polymerase active center. *Science*, 273:107-9.
- Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR. 2000. Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J. Biol. Chem.*, 275:8103-13.
- Zeppenfeld T, Larisch C, Lengeler JW, Jahreis K. 2000. Glucose transporter mutants of *Escherichia coli* K-12 with changes in substrate recognition of IICB^{Glc} and induction behavior of the *ptsG* gene. *J. Bacteriol.*, 182:4443-52.
- Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G. 1998a. The *OxyS* regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J.*, 17:6061-8.
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, 98:811-24.
- Zhang XP, Ebright RH. 1990. Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base pair 5 of the DNA site in the CAP-DNA complex. *Proc. Natl. Acad. Sci. USA*, 87:4717-21.
- Zhao G, Pease AJ, Bharani N, Winkler ME. 1995. Biochemical characterization of *gapB*-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and its possible role in pyridoxal 5'-phosphate biosynthesis. *J. Bacteriol.*, 177:2804-12.
- Zhao R, Collins EJ, Bourret RB, Silversmith RE. 2002. Structure and catalytic mechanism of the

Chez *Escherichia coli*, la glyceraldéhyde 3-phosphate déshydrogénase (GAPDH), qui est une enzyme clé du métabolisme du glucose est codée par le gène *gapA*. La transcription de ce gène est initiée au niveau de 4 promoteurs et majoritairement au niveau du promoteur *gapA* P1 en phase exponentielle de croissance. Nous avons montré qu'il a la particularité d'avoir une boîte -10 étendue, mais de nécessiter la présence d'une boîte -35 pour être efficace. La méthionine sulfoxyde réductase de type B (MsrB) est codée par le gène *yeaA* situé en amont et en sens opposé à *gapA*. Nous avons montré que des éléments de séquences inter-ORF régulent chacun des deux gènes soit dans le même sens, soit en sens opposé et nous avons obtenu des arguments en faveur d'une régulation traductionnelle de l'expression de la protéine MsrB par le petit ARN non codant RyhB. Nous avons développé l'approche puces à ADN pour tester l'effet de l'inactivation du transporteur EIIBC^{Glc} sur le transcriptome d'*E. coli*. Nous avons ainsi montré que l'expression de gènes permettant la formation de pores de plus grande taille dans la membrane externe et de flagelles favorisant la motilité de la bactérie était augmentée.

Mots clés : *Escherichia coli*, glyceraldéhyde 3-phosphate déshydrogénase (GAPDH), initiation de la transcription, promoteur étendu, transcriptome, méthionine sulfoxyde réductase de type B (MsrB), ARNnc RyhB, porines, flagelles.

Title : Study on the regulation of genes expression in *Escherichia coli* in relation with transport and glucose metabolism, including :

- a molecular analysis of the promoter region which directs expression of the *gapA* and *yeaA* genes,
- a transcriptomic study on the effect of the absence of the EIIBC^{Glc} protein required for the specific transport of glucose in *Escherichia coli*.

Summary :

In *Escherichia coli*, the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is a key enzyme for glucose metabolism and is encoded by the *gapA* gene. The transcription of this gene is initiated at 4 promoters and mainly at the *gapA* P1 promoter in the exponential phase of growth. We found that although being a promoter of the -10 extended class, promoter P1 needs a -35 hexamer for activity. The *E. coli* methionine sulfoxide reductase of the B type (MsrB) is encoded by the *yeaA* gene which is located upstream and in the opposite direction of *gapA*. We showed that elements of the *gapA-yeaA* inter-ORF sequence are involved in the regulation of the two genes, either activating both genes or having opposite effects on them and we obtained data in favor of a translation regulation by the small non-coding RNA RyhB of the MsrB protein expression. We developed a DNA macroarray approach to test for the effect of the inactivation of the EIIBC^{Glc} protein on the *E. coli* transcriptome. We showed that the expression of genes allowing formation of large pores in the external membrane and flagella were upregulated. We demonstrated the increased motility of the bacteria.

Key words : *Escherichia coli*, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), transcription initiation, extended promoter, transcriptome, methionine sulfoxide reductase of the B type (MsrB), sRNA RyhB, porins, flagella.