



HAL
open science

Les Glyceraldehyde-3-phosphate deshydrogenases de deux bactéries : Clostridium pasteurianum et bacillus stearothermophilus : séquence de leur gène et expression chez escherichia coli : étude phylogénétique de la glyceraldehyde-3-phosphate deshydrogénase

Thierry Oster

► **To cite this version:**

Thierry Oster. Les Glyceraldehyde-3-phosphate deshydrogenases de deux bactéries : Clostridium pasteurianum et bacillus stearothermophilus : séquence de leur gène et expression chez escherichia coli : étude phylogénétique de la glyceraldehyde-3-phosphate deshydrogénase. Biochimie, Biologie Moléculaire. Université Henri Poincaré - Nancy 1, 1989. Français. NNT : 1989NAN10023 . tel-01748139

HAL Id: tel-01748139

<https://hal.univ-lorraine.fr/tel-01748139v1>

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>

THESE

présentée à l'U.E.R. PHYSIQUE - CHIMIE - BIOLOGIE

pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE DE NANCY I
en BIOLOGIE MOLECULAIRE et GENIE GENETIQUE**

par

Thierry OSTER



**Les glycéraldéhyde-3-phosphate déshydrogénases de deux bactéries :
Clostridium pasteurianum et *Bacillus stearothermophilus*.
Séquence de leur gène et expression chez *Escherichia coli*.**

Etude phylogénétique de la glycéraldéhyde-3-phosphate déshydrogénase.

Soutenue publiquement le 30 janvier 1989 devant la Commission d'Examen :

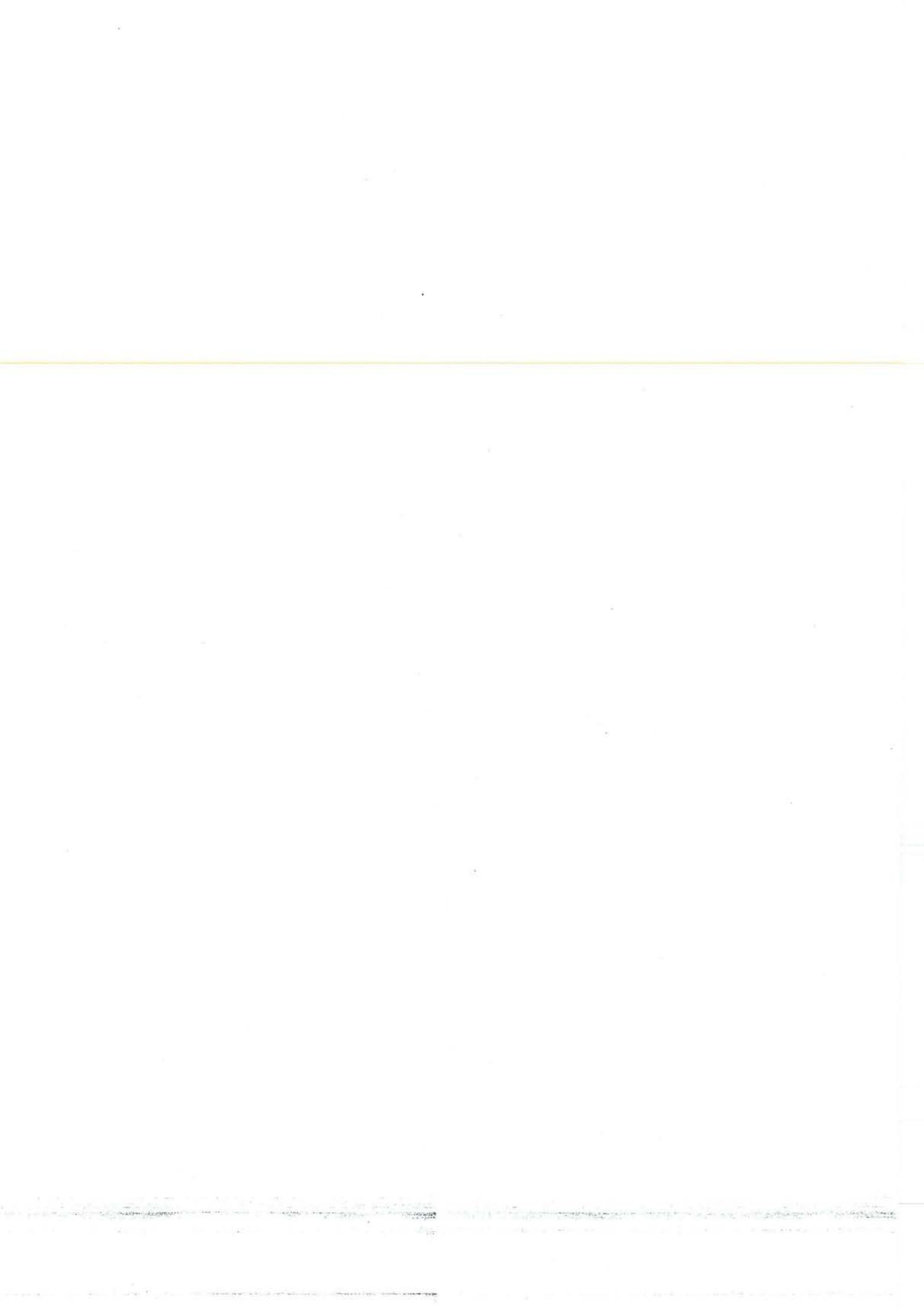
M. le Pr. G. LINDEN	Président de Jury ; Nancy
M. le Pr. R. CERFF	Rapporteur externe ; Grenoble
M. le Pr. H. PETITDEMANGE	Rapporteur interne ; Nancy
Mme le Dr. C. BRANLANT	Directeur de Recherche au CNRS ; Nancy
M. le Pr. G. BRANLANT	Laboratoire d'Enzymologie et Génie Génétique ; Nancy
M. le Pr. B. EHRESMANN	Institut de Biologie Moléculaire et Cellulaire ; Strasbourg
M. le Pr. G. SIEST	Centre du Médicament ; Nancy

à Isabelle,

à mes parents.



Ce travail a été réalisé au Laboratoire d'Enzymologie et de Génie Génétique (Unité Associée au CNRS n° 457) de la Faculté des Sciences de l'Université de Nancy I, dirigé par le Professeur G. BRANLANT.



AVANT-PROPOS

J'exprime ma profonde reconnaissance à Monsieur le Professeur Guy BRANLANT pour m'avoir accueilli dans son laboratoire et suivi avec intérêt l'évolution de mes travaux.

J'adresse mes plus vifs remerciements à Madame le Docteur Christiane BRANLANT qui a dirigé ce travail. Sa disponibilité, ses conseils très précieux et sa rigueur scientifique m'ont permis de mener à bien ce travail et m'ont été indispensables jusque dans la rédaction de cette thèse.

Je tiens à témoigner de ma profonde gratitude à Monsieur le Professeur Guy LINDEN qui a bien voulu me faire l'honneur de présider le jury de cette thèse.

Je remercie Messieurs les Professeurs Rüdiger CERFF et Henri PETITDEMANGE qui ont accepté de juger ce travail.

Mes remerciements s'adressent également à Messieurs les Professeurs Bernard EHRESMANN et Gérard SIEST pour avoir bien voulu participer aux travaux de ce jury.

J'adresse mes remerciements à Mesdames Jacqueline BAYEUL et Evelyne HABERMACHER pour leur assistance technique précieuse et de qualité, ainsi qu'à Monsieur Jean-Paul DECLE pour l'enthousiasme et l'application de son soutien technique.

Je remercie également tous mes camarades de travail pour leur bonne humeur et les bons moments passés ensemble, pour leur soutien et leur gentillesse. Je pense en particulier à Monsieur Omar ASSOBBHEI qui a collaboré à une partie de ce travail.

Je tiens à remercier enfin Isabelle, Laurence et Martine pour leur assistance dactylographique efficace et dévouée, ainsi que Colette et Marcel pour leur patience et la qualité de leurs corrections d'orthographe et de syntaxe.

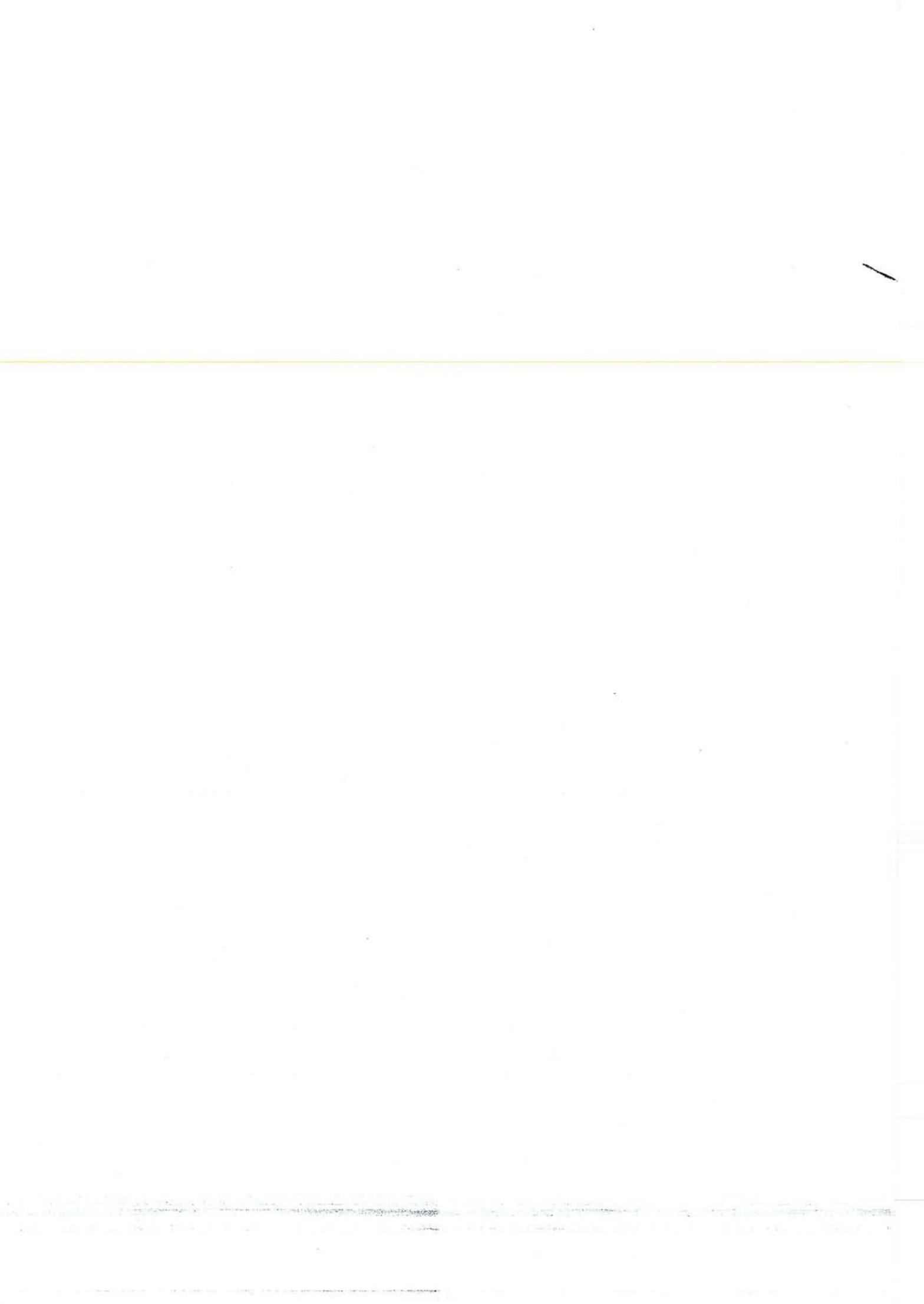


TABLE DES MATIERES

ABREVIATIONS.....	page	2
ILLUSTRATIONS ET TABLEAUX.....	page	4
CONSTRUCTIONS PLASMIDIQUES.....	page	6
INTRODUCTION.....	page	7
MATERIELS ET METHODES.....	page	43
RESULTATS ET DISCUSSIONS.....	page	99
BILAN ET PERSPECTIVES.....	page	215
REFERENCES BIBLIOGRAPHIQUES.....	page	221
RESUME.....	page	237

ABBREVIATIONS

aa	: acide(s) aminé(s)
cAMP	: AMP cyclique
ATCC	: "American Type Culture Collection"
ATP	: adénosine triphosphate
β -ME	: β -mercapto-éthanol
Bls	: Bluescript
BrEt	: bromure d'éthidium
<i>B.stearothermophilus</i>	: <i>Bacillus stearothermophilus</i>
<i>C.pasteurianum</i>	: <i>Clostridium pasteurianum</i>
db, sb	: double brin, simple brin
DMTr	: diméthoxytrityl
(c)DNA	: acide(s) déoxyribonucléique(s) (complémentaire)
(d)dNTP	: (di)déoxynucléotide(s) triphosphate
DO	: densité optique
1,3 DPG	: 1,3 diphosphoglycérate
DTT	: dithiothréitol
<i>E.coli</i>	: <i>Escherichia coli</i>
EDTA	: éthylène diamine tétra-acétate
GAP	: glycéraldéhyde-3-phosphate
GAPDH	: glycéraldéhyde-3-phosphate deshydrogénase(s)
HPLC	: chromatographie liquide à haute pression
IPTG	: isopropyl β -D-thiogalactoside
lac	: lactose
LB	: Luria "broth"
mer	: oligomère
<i>M.fervidus</i>	: <i>Methanothermus fervidus</i>

NAD(P)	: nicotinamide adénine dinucléotide (phosphate)
NCA	: "National Canner's American"
ORF	: phase(s) ouverte(s) de lecture ("open reading frame")
pb, kb	: paire(s) de bases, kilobase(s)
PEG	: polyéthylène-glycol
3 PG	: 3 phosphoglycérate
PGK	: phosphoglycérate kinase
Pi	: phosphate inorganique
promoteur bla	: promoteur du gène de la β -lactamase
promoteur gap	: promoteur du gène de la GAPDH d' <i>E.coli</i>
p/v	: poids/volume
RBS	: site de fixation du ribosome ("ribosome binding site")
(m,r,t)RNA	: acide(s) ribonucléique(s) (messenger, ribosomal, de transfert)
<i>S.cerevisiae</i>	: <i>Saccharomyces cerevisiae</i>
SD	: Shine-Dalgarno
SDS	: sodium dodécylsulfate
<i>S.lactis</i>	: <i>Streptococcus lactis</i>
<i>T.aquaticus</i>	: <i>Thermus aquaticus</i>
TE	: Tris-EDTA
TEA	: Tris-EDTA-acétate
TEAA	: acétate de triéthylammonium
TEB	: Tris-EDTA-borate
TEMED	: N, N, N', N' tétraméthyl éthylène diamine
Tris	: Tris (hydroxyméthyl) aminométhane
<i>T.tenax</i>	: <i>Thermoproteus tenax</i>
U.V.	: Ultra-Violets
v/v, vol/vol	: volume/volume
X-Gal	: 5-bromo, 4-chloro, 3-indolyl β -D-galactopyranoside
<i>Z.mobilis</i>	: <i>Zymomonas mobilis</i>
<i>Z.rouxii</i>	: <i>Zygosaccharomyces rouxii</i>

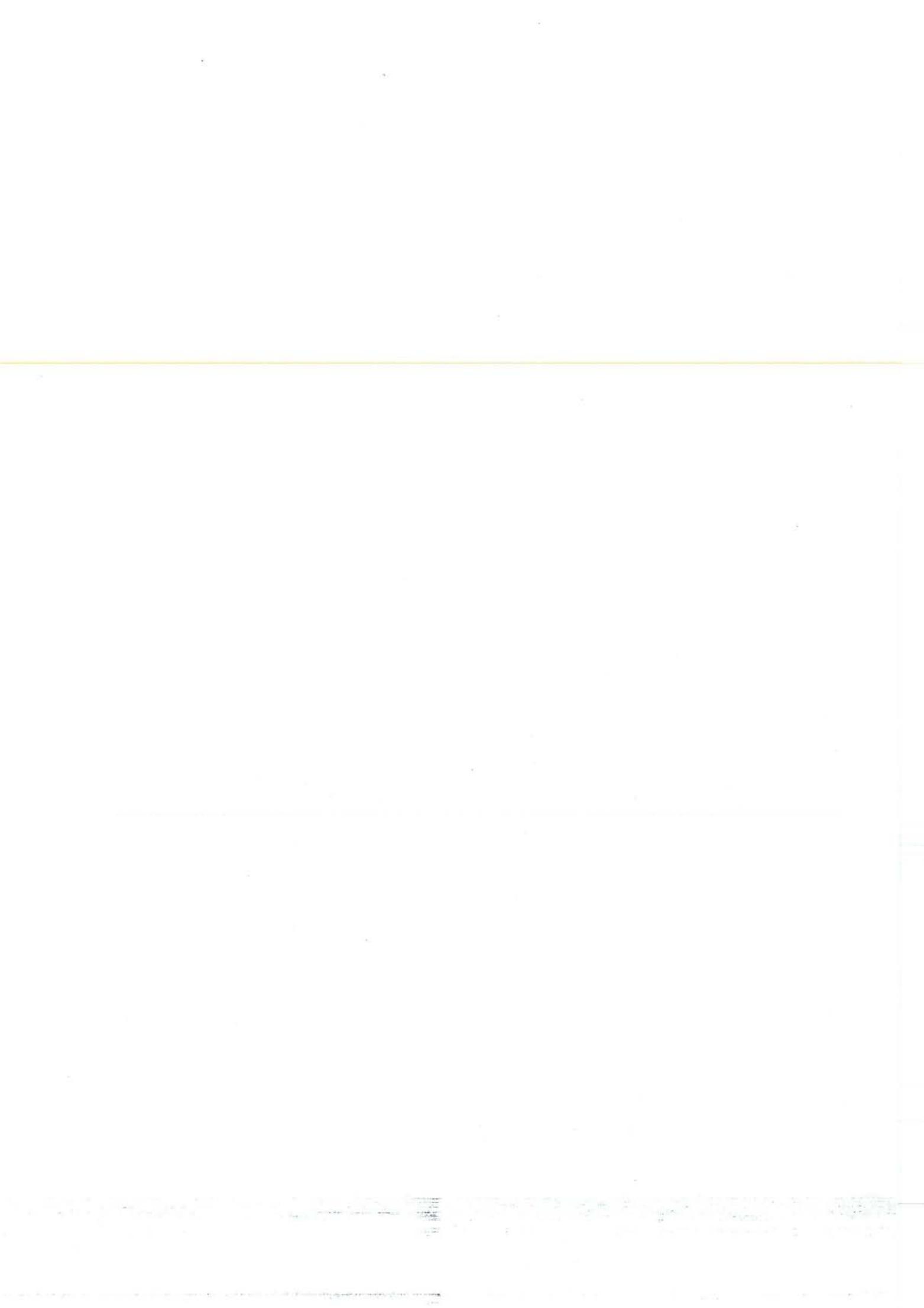
ILLUSTRATIONS ET TABLEAUX

Fig. 1	: La voie glycolytique.....	page 13
Fig. 2	: Le cycle de fixation du CO ₂ de Calvin-Benson.....	page 15
Fig. 3	: Mécanisme de transfert indirect du NADP du chloroplaste au cytosol.....	page 17
Fig. 4	: Structure tridimensionnelle du monomère de la GAPDH de <i>B.stearothermophilus</i>	page 20
Fig. 5	: Formation de l'intermédiaire thiohémicétal.....	page 23
Fig. 6	: Formation de l'acyl-enzyme.....	page 25
Fig. 7	: Formation de l'intermédiaire phosphorylé.....	page 26
Fig. 8	: Vue stéréo du site actif de la GAPDH de <i>B.stearothermophilus</i>	page 27
Fig. 9	: Fixation du glycidol-3-phosphate dans le site actif de la forme holoenzyme de la GAPDH de <i>B.stearothermophilus</i>	page 29
Fig. 10	: Les sites anioniques et le site de fixation du NAD de la GAPDH de <i>B.stearothermophilus</i>	page 31
Fig. 11	: Dégradation du glucose et du xylose chez <i>C.pasteurianum</i>	page 38
Fig. 12	: La réaction phosphoroclastique chez <i>C.pasteurianum</i>	page 40
Fig. 13	: Principales voies de dégradation de l'acétyl-CoA chez <i>C.pasteurianum</i>	page 41
Fig. 14	: Cartographie du plasmide pBR322.....	page 54
Fig. 15	: Cartographie des plasmides pUC.....	page 56
Fig. 16	: Cartographie des plasmides Bluescript.....	page 57
Fig. 17	: Le phage M13 et ses dérivés.....	page 58
Fig. 18	: Le cycle de vie du phage M13.....	page 65
Fig. 19	: Synthèse chimique des oligonucléotides. Méthode aux phosphites.....	page 74
Fig. 20	: Structure d'un déoxynucléotide et d'un didéoxynucléotide.....	page 77
Fig. 21	: Principe du séquençage par la méthode aux didéoxynucléotides de Sanger.....	page 77
Fig. 22	: Principe de la méthode de mutagenèse dirigée utilisée.....	page 85
Fig. 23	: Cinétiques d'hydrolyse du DNA génomique de <i>C.pasteurianum</i>	page 105
Fig. 24	: Fractionnement en gel d'agarose des produits d'hydrolyse de pCGap1.....	page 110
Fig. 25	: Cartographie des principaux sites de restriction de pCGap1.....	page 111
Fig. 26	: Construction des plasmides pCGap2 et pCGap3.....	page 113
Fig. 27	: Fractionnement en gel d'agarose des produits d'hydrolyse de pCGap2 et pCGap3.....	page 114
Fig. 28	: Sous-clonage du fragment DraI de 1,8 kb.....	page 116
Fig. 29	: Stratégie du séquençage de pCGap1.....	page 118
Fig. 30	: Séquence nucléotidique du fragment de 3.663 pb cloné dans pCGap1 et contenant le gène de la GAPDH de <i>C.pasteurianum</i>	page 121

Fig. 31	: Les différentes ORF du fragment cloné dans pCGap1.....	page 127
Fig. 32	: Séquence nucléotidique de la phase ouverte G3 et séquence de la protéine correspondante.....	page 129
Fig. 33	: Séquence nucléotidique de la phase ouverte G4 et séquence de la protéine correspondante.....	page 130
Fig. 34	: Séquence nucléotidique de la phase ouverte G5 et séquence de la protéine correspondante.....	page 131
Fig. 35	: Comparaison des séquences promotrices potentielles des gènes de <i>C.pasteurianum</i>	page 132
Fig. 36	: Signaux potentiels de terminaison de la transcription Rho-indépendants en aval des gènes de <i>C.pasteurianum</i>	page 134
Fig. 37	: Les séquences RBS potentielles des gènes de <i>C.pasteurianum</i> contenus dans pCGap1.....	page 136
Tab. I	: Usage des codons dans les gènes de GAPDH de <i>C.pasteurianum</i> et d' <i>E.coli</i>	page 139
Fig. 38	: Construction du plasmide pCGap5.....	page 148
Fig. 39	: Carte de restriction du vecteur d'expression Bls-gap.....	page 150
Fig. 40	: Construction du plasmide pCGap6.....	page 152
Fig. 41	: Fractionnement en gel d'agarose des produits d'hydrolyse de pCGap6.....	page 153
Tab. II	: Activité spécifique de la GAPDH dans les différentes souches recombinantes DF221.....	page 155
Fig. 42	: Fractionnement en gel d'acrylamide des extraits cellulaires des différentes souches recombinantes DF221.....	page 156
Fig. 43	: Localisation des fragments où persistaient des ambiguïtés de séquence au sein du fragment génomique de 4.038 pb cloné dans pBst-gap1.....	page 161
Fig. 44	: Séquence nucléotidique du fragment de 4.038 pb cloné dans pBst-gap1 et contenant le gène de la GAPDH de <i>B.stearothermophilus</i>	page 163
Fig. 45	: Carte de restriction du plasmide pBst-gap1.....	page 167
Fig. 46	: Hybridations par la méthode de Southern des sous-fragments issus du fragment de 4.038 pb sur le DNA génomique de <i>B.stearothermophilus</i>	page 169
Fig. 47	: Les séquences RBS potentielles des gènes de <i>B.stearothermophilus</i> contenus dans pBst-gap1...	page 184
Fig. 48	: Localisation des différentes séquences promotrices potentielles en amont du gène de la GAPDH de <i>B.stearothermophilus</i>	page 187
Fig. 49	: Alignement des structures primaires de GAPDH de différents organismes.....	page 196
Fig. 50	: Localisation des régions conservées et divergentes dans la structure tridimensionnelle de la GAPDH.....	page 202
Fig. 51	: Les régions conservées dans la structure tétramérique de la GAPDH.....	page 202
Tab. III	: Pourcentages d'homologie totale entre les GAPDH.....	page 205
Tab. IV	: Pourcentages d'homologie entre les domaines de fixation du coenzyme des GAPDH.....	page 207
Tab. V	: Pourcentages d'homologie entre les domaines catalytiques des GAPDH.....	page 208
Fig. 52	: Alignement des séquences de "S-loop" des GAPDH.....	page 210
Tab. VI	: Pourcentages d'homologie entre les "S-loop" des GAPDH.....	page 211

CONSTRUCTIONS PLASMIDIQUES

- pCGap1** : plasmide pBR322 contenant le fragment génomique de 3.663 pb renfermant le gène de la GAPDH de *C.pasteurianum* (Fig. 25)
- pCGap2** : plasmide obtenu après élimination du fragment HindIII de pCGap1 (Fig. 26)
- pCGap3** : plasmide obtenu après élimination du fragment EcoRI de pCGap1 (Fig. 26)
-
- pCGap4** : plasmide pUC19 contenant le fragment DraI dans lequel le gène de la GAPDH de *C.pasteurianum* est orienté en sens inverse du promoteur lacZ (Fig. 28)
- pCGap5** : plasmide pUC18 contenant le fragment DraI dans lequel le gène de la GAPDH de *C.pasteurianum* est orienté dans le même sens que le promoteur lacZ (Fig. 38)
- pCGap6** : plasmide Bls⁺ contenant la partie codante du gène de la GAPDH de *C.pasteurianum* sous le contrôle du promoteur gap d'*E.coli* (Fig. 40)
- pBst-gap1** : plasmide pBR322 contenant le fragment génomique de 4.038 pb renfermant le gène de la GAPDH de *B.stearothermophilus* (Fig. 45)
- pBst-gap2** : plasmide pBR322 contenant le fragment HindIII de pBst-gap1 (Fig. 3 de la publication)
- pBst-gap3** : plasmide obtenu après élimination du fragment BamHI-SmaI de pBst-gap1 (Fig. 3 de la publication)
- pBst-gap4** : plasmide pBR322 recombinant, isolé en même temps que pBst-gap1, contenant un fragment génomique plus court que celui cloné dans pBst-gap1 et renfermant le gène de la GAPDH de *B.stearothermophilus*
- pF61a** : plasmide pBR322 contenant un fragment génomique renfermant le gène de la GAPDH d'*E.coli*



INTRODUCTION

Les glycéraldéhyde-3-phosphate déshydrogénases (GAPDH) peuvent être subdivisées en trois grands types d'enzymes caractérisés par des localisations et des rôles biologiques différents.

Parmi elles, la GAPDH glycolytique fait partie des enzymes les plus étudiées jusqu'à ce jour, en raison de son importance physiologique, mais aussi de sa facilité d'isolement. Elle a été étudiée essentiellement selon trois approches :

- identification des résidus essentiels dans la catalyse enzymatique, par les techniques classiques d'enzymologie ;
- études cristallographiques permettant de proposer un modèle de structure tertiaire et d'identifier les amino-acides susceptibles d'être impliqués dans la fixation du substrat et du cofacteur ;
- plus récemment, au laboratoire, étude fine du site actif par la technique de mutagenèse dirigée.

La GAPDH glycolytique présente une forte conservation structurale au cours de l'évolution, et ceci chez des organismes aussi éloignés que les bactéries, les levures, les animaux et les végétaux, ce qui fait de cette enzyme un bon outil d'étude des problèmes d'évolution.

La mise en évidence de deux autres types de GAPDH a suscité également le développement d'études phylogénétiques approfondies.

SOMMAIRE

I - LES DIFFERENTS TYPES DE GAPDH

1. GAPDH impliquée dans la glycolyse et la gluconéogénèse.....	page 11
1-1. Localisation.....	page 11
1-2. Rôle métabolique.....	page 12
2. GAPDH intervenant dans le cycle de Calvin-Benson.....	page 14
2-1. Localisation.....	page 14
2-2. Rôle métabolique.....	page 14
3. GAPDH dépourvue d'activité phosphorylante.....	page 14
3-1. Localisation.....	page 16
3-2. Rôle métabolique.....	page 16

II - ETUDES STRUCTURALES ET ENZYMOLOGIQUES DE LA GAPDH GLYCOLYTIQUE

1. Structure des GAPDH.....	page 18
1-1. Structure primaire.....	page 18
1-2. Structure tertiaire.....	page 19
1-2-1. Le domaine de fixation du coenzyme.....	page 19
1-2-2. Le domaine catalytique.....	page 21
1-3. Symétrie moléculaire.....	page 21
1-4. Relations structure-stabilité.....	page 21
2. Mécanisme enzymatique de la GAPDH glycolytique.....	page 22
2-1. Formation de l'intermédiaire thiohémicétal.....	page 22
2-2. Formation de l'acyl-enzyme.....	page 24
2-3. Formation de l'intermédiaire phosphorylé.....	page 24
2-4. Les amino-acides impliqués dans ce mécanisme.....	page 24
2-4-1. Ceux intervenant dans la catalyse enzymatique.....	page 24
2-4-2. Ceux impliqués dans la fixation du GAP et du Pi.....	page 28
2-4-3. Ceux impliqués dans la fixation du NAD et ceux responsables de la spécificité de cofacteur.....	page 28
3. Etude de l'allostérie.....	page 32

III - LES GENES DE GAPDH

1. Les gènes de GAPDH du règne animal.....	page 33
2. Les gènes de GAPDH du règne végétal.....	page 34
3. Les gènes de GAPDH de levures.....	page 35
4. Les gènes de GAPDH bactériennes.....	page 35

IV - LES MICRO-ORGANISMES CHOISIS

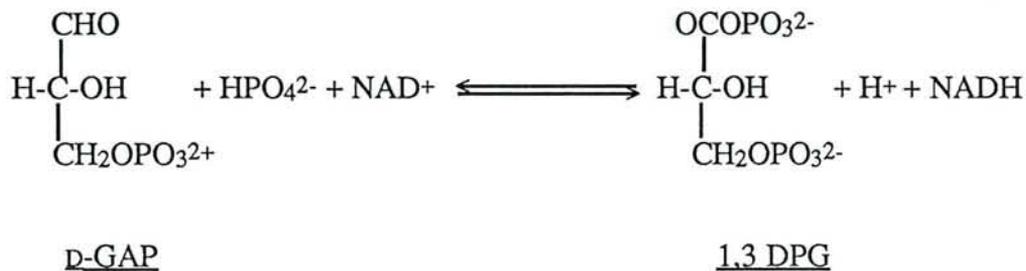
1. Les signaux de régulation des gènes de bactéries Gram-positif.....	page 36
2. <i>Clostridium pasteurianum</i>	page 37
3. <i>Bacillus stearothermophilus</i>	page 39

I - LES DIFFERENTS TYPES DE GAPDH

Les premières GAPDH étudiées étaient celles intervenant dans la glycolyse. Des travaux plus récents ont permis de caractériser deux autres types de GAPDH ayant un rôle biologique différent.

1. GAPDH impliquée dans la glycolyse et la gluconéogénèse (EC 1-2-1-12)

Elle catalyse réversiblement la phosphorylation oxydative du D-glycéraldéhyde-3-phosphate (GAP) en acide 1,3 diphosphoglycérique (1,3 DPG). La réaction enzymatique se décompose en deux étapes : une étape d'oxydo-réduction, conduisant à un intermédiaire acyl-enzyme, et une étape de phosphorylation, conduisant au 1,3 DPG. Le schéma global de la réaction est le suivant :



1-1. Localisation

La GAPDH glycolytique est présente aussi bien au niveau des cellules procaryotiques qu'eucaryotiques. Les quantités de GAPDH sont très importantes chez les micro-organismes et dans les cellules eucaryotiques spécialisées dans la glycolyse tels les muscles.

Chez *Zymomonas mobilis*, bactérie fermentaire à très forte activité glycolytique, les enzymes de la glycolyse représentent jusqu'à 50 % des protéines solubles dont 1/10^e de GAPDH (Algar et Scopes, 1985 ; Pawluk et al., 1986). La GAPDH représente plus de 10 % des protéines solubles dans le muscle de lapin (Cori et al., 1945).

Une enzyme de ce type a récemment été isolée à partir de glycosomes (Misset et al., 1987 ; Read et al., 1987).

1-2. Rôle métabolique

La voie glycolytique (chemin de Embden-Meyerhoff-Parnas) conduit à la conversion du glucose en acide pyruvique. Dans la plupart des organismes, cette voie est essentielle au métabolisme des glucides, notamment celui du glucose (Fig. 1).

La réaction catalysée par la GAPDH dans cette voie métabolique est particulièrement importante puisqu'elle permet de coupler une déshydrogénation et une phosphorylation, et ainsi de stocker l'énergie d'oxydation dans un acyl-phosphate.

Dans l'étape suivante de la glycolyse, la phosphoglycérate kinase, l'enzyme juste en aval de la GAPDH, catalyse la réaction utilisant le haut potentiel de transfert du groupement phosphate du 1,3 DPG pour produire de l'ATP. Cette réaction constitue la première étape génératrice d'ATP de la glycolyse.

Chez les eucaryotes, la GAPDH est impliquée à la fois dans la voie de la glycolyse et celle de la gluconéogénèse. La réversibilité de la réaction qu'elle catalyse dépend de l'état d'oxydation des cellules, c'est-à-dire du rapport NAD^+/NADH régnant dans le cytosol. Une diminution de ce rapport, consécutive à un régime hypoglucidique par exemple, favorisera la gluconéogénèse. D'autre part, dans le cas de cellules cancéreuses en croissance nécessitant d'importantes réserves énergétiques, il a été montré que le gène de la GAPDH était fortement surexprimé (Tokunaga et al., 1987).

Shonk et Boxer (1964) ont suggéré l'intervention de la GAPDH dans la régulation de la glycolyse dans différents tissus tels que le foie, les muscles, le cerveau, le coeur ou le tissu adipeux. Chez la levure, il semble que la phosphoglycérate kinase joue un rôle essentiel dans cette régulation en association avec la GAPDH (Ashmarina et al., 1985).

Cette notion de régulation de la glycolyse par la GAPDH a été récemment étendue à certains procaryotes. Chez *Streptococcus lactis* et *Streptococcus cremoris*, la diminution de la glycolyse observée dans les cellules cultivées en l'absence de lactose serait liée à une diminution de l'activité de la GAPDH (Poolman et al., 1987).

Enfin, Kawamoto et Caswell (1986) ont montré que la GAPDH du muscle de lapin pouvait participer à des fonctions indépendantes de la glycolyse. Cette enzyme serait capable de s'auto-phosphoryler en présence de Mg^{2+} et d'ATP. La protéine phosphorylée se comporterait alors comme une kinase capable de catalyser la phosphorylation des protéines membranaires des microsomes de muscle squelettique.

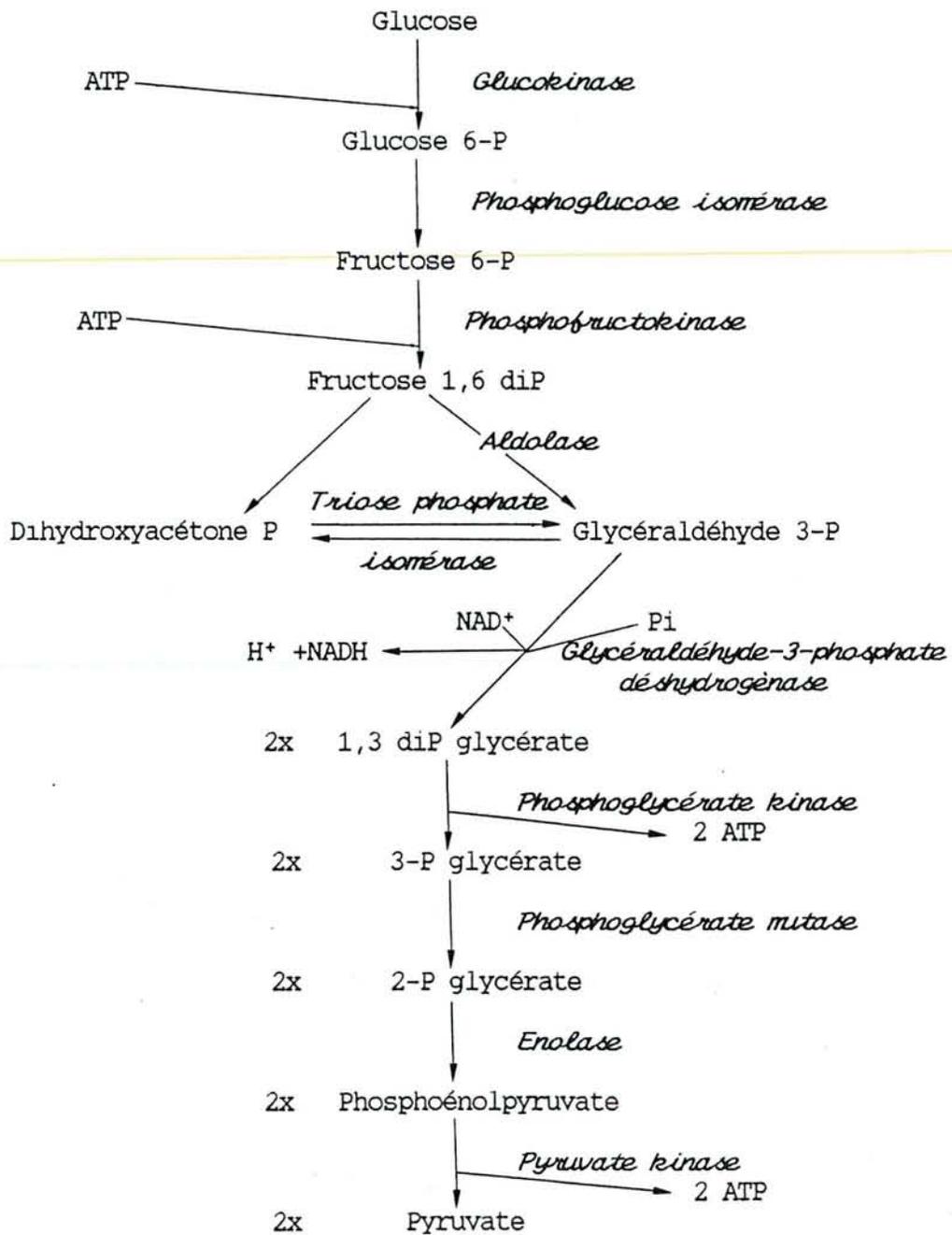
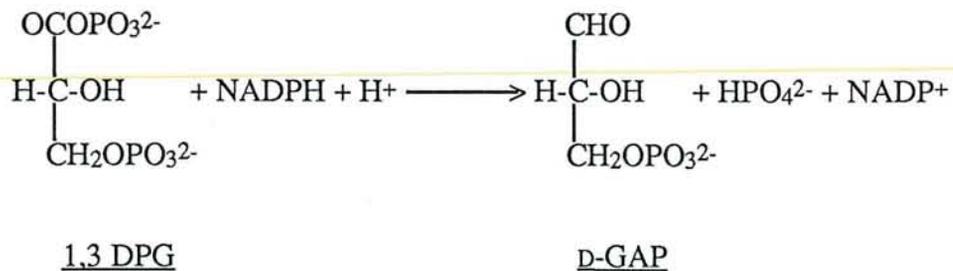


Figure 1 : La voie glycolytique

2. GAPDH intervenant dans le cycle de Calvin-Benson (EC 1-2-1-13)

Ce type de GAPDH catalyse la réduction et la déphosphorylation irréversible du 1,3 diphosphoglycérate en glycéraldéhyde-3-phosphate, en utilisant préférentiellement le NADP⁺ comme cofacteur (Mc Gowan et Gibbs, 1974 ; Martin et Cerff, 1986). Le schéma de la réaction est le suivant :



La GAPDH chloroplastique des végétaux supérieurs consiste en deux isoenzymes tétramériques (Cerff, 1979) différant légèrement par le poids moléculaire des sous-unités (A < B) qui les composent, A₂B₂ ou A₄ (Cerff et Chambers, 1979).

2-1. Localisation

Ce type de GAPDH a été identifié dans les chloroplastes des végétaux chlorophylliens.

2-2. Rôle métabolique

Cette GAPDH intervient dans le cycle de Calvin-Benson (Fig. 2). Elle participe aux mécanismes de fixation du CO₂ lors de la photosynthèse, permettant la production de glucides, acides gras et acides aminés chez les végétaux chlorophylliens. Comme de nombreuses enzymes de ce cycle, la GAPDH chloroplastique est photo-inductible (Cerff et Kloppstech, 1982; Shih et Goodman, 1988).

3. GAPDH dépourvue d'activité phosphorylante (EC 1-2-1-9)

Ce type de GAPDH catalyse l'oxydation irréversible du GAP en 3 PG en utilisant exclusivement le NADP⁺ comme cofacteur (Iglesias et al., 1987 ; Iglesias et Losada, 1988).

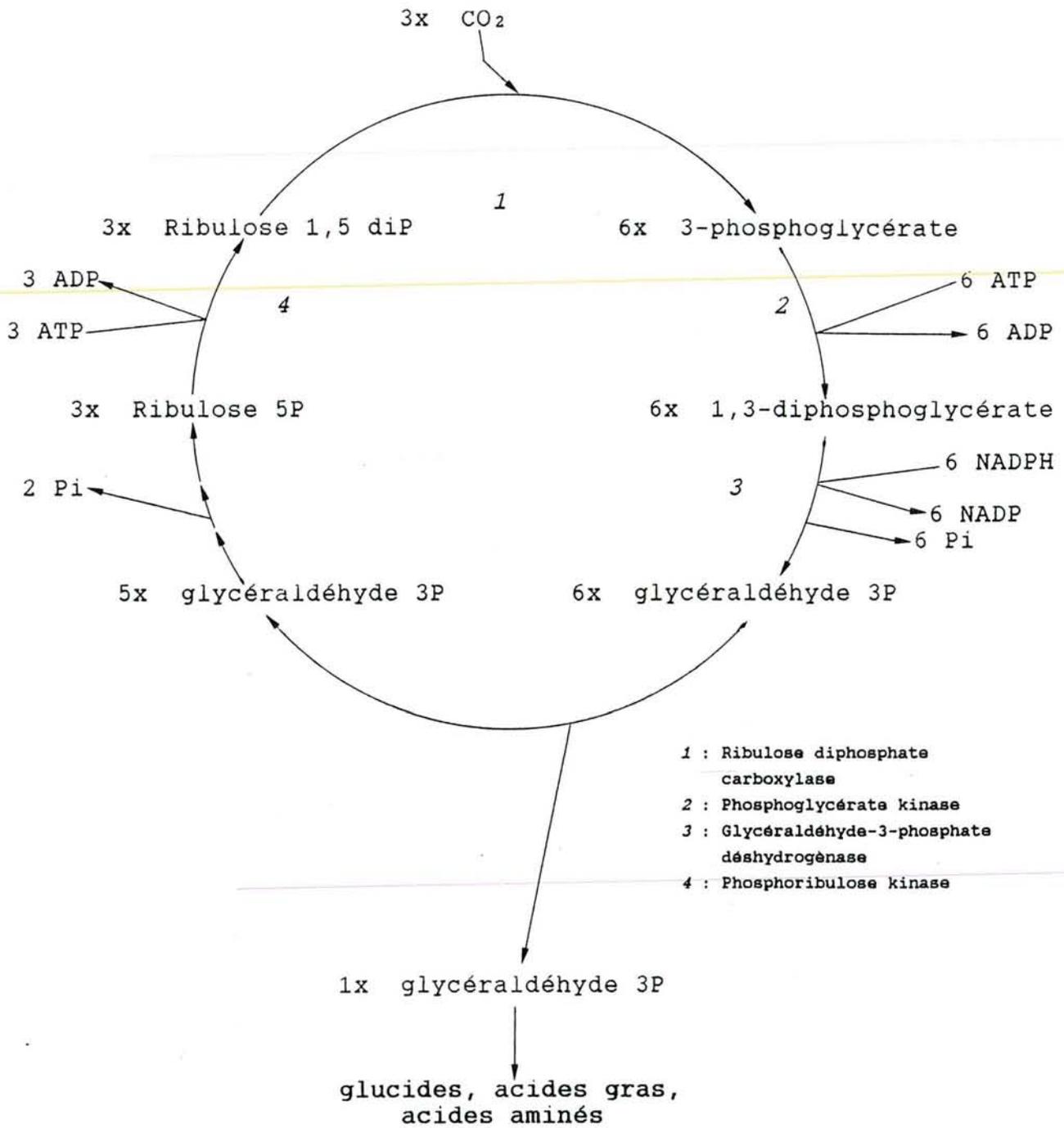
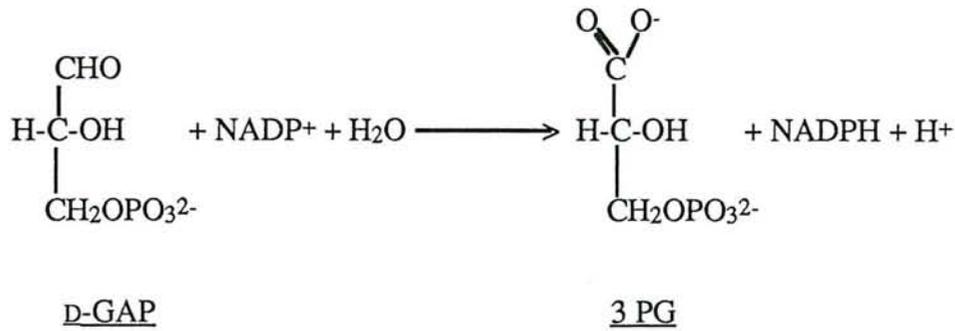


Figure 2 : Le cycle de fixation du CO₂ de Calvin-Benson

Le schéma global de la réaction est le suivant :



Cette enzyme se présente sous la forme d'un homotétramère, d'un poids moléculaire d'environ 195.000 D chez l'épinard (Iglesias et Losada, 1988).

3-1. Localisation

Ce type de GAPDH a été isolé à partir du cytoplasme de végétaux chlorophylliens supérieurs et d'une algue verte, *Chlamydomonas reinhardtii* (Iglesias et al., 1987).

3-2. Rôle métabolique

Bamberger et al. (1975) ont suggéré l'implication possible de cette enzyme dans la création d'un gradient de protons de part et d'autre de la membrane chloroplastique. Kelly et Gibbs (1973) pensent qu'elle participerait, couplée à la GAPDH chloroplastique, au mécanisme de transfert indirect du NADPH du chloroplaste vers le cytoplasme. Cette enzyme, activée par le NADP⁺ cytosolique, interviendrait donc dans un système régulant l'équilibre entre NADP et NADPH (Fig. 3).

Remarque : Des travaux récents sur des archaebactéries thermophiles ont permis de mettre en évidence de nouvelles formes de GAPDH. Ainsi, chez *Methanothermus fervidus*, une GAPDH glycolytique stable jusqu'à 75°C a été isolée. D'un poids moléculaire d'environ 200.000 D, cet homotétramère présenterait la particularité de fonctionner avec le NAD⁺ ou le NADP⁺ comme cofacteur (Fabry et Hensel, 1987).

Hensel et al. (1987) ont, par ailleurs, caractérisé deux GAPDH chez *Thermoproteus tenax*. Ces deux enzymes, l'une NAD⁺ spécifique, l'autre NADP⁺ spécifique, sont stables jusqu'à 96°C et existent en proportions variables selon les conditions de croissance.

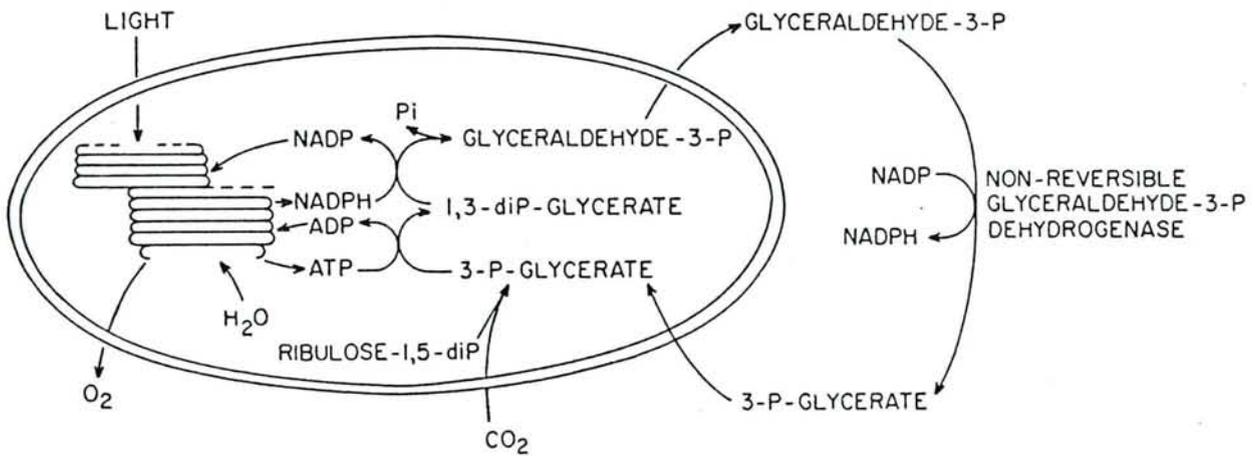


Figure 3 : Mécanisme de transfert indirect du NADP du chloroplaste vers le cytosol, d'après Kelly et Gibbs (1973)

II - ETUDES STRUCTURALES ET ENZYMOLOGIQUES DE LA GAPDH GLYCOLYTIQUE

1. Structure des GAPDH

La GAPDH glycolytique est une enzyme tétramérique formée de sous-unités identiques d'un poids moléculaire de 36.000 D et d'environ 330 acides aminés chacun.

1-1. Structure primaire

Une vingtaine de séquences primaires de GAPDH ont été déterminées à partir d'espèces parfois très éloignées dans l'évolution.

Parmi les enzymes de procaryotes, les séquences primaires des GAPDH de *Thermus aquaticus* et de *Bacillus stearothermophilus* ont été établies directement à partir de la protéine (Hocking et Harris, 1980 ; Biesecker et al., 1977 ; Walker et al., 1980a). Par contre, la séquence primaire d'*Escherichia coli* a été déduite de la séquence nucléotidique du gène correspondant (Branlant et Branlant, 1985). Il s'agissait là des seules séquences en acides aminés connues au début de notre travail. Depuis, la structure primaire de la GAPDH de *Zymomonas mobilis* (Conway et al., 1987) a été déterminée à partir de son gène. Fabry et Hensel (1988) ont déterminé la séquence d'un gène qui coderait pour l'enzyme qu'ils avaient identifiée comme étant la GAPDH de *Methanothermus fervidus* (Fabry et Hensel, 1987). De toute évidence, d'après la structure primaire de la protéine déduite de la séquence nucléotidique de ce gène, comparativement à l'ensemble des séquences de GAPDH connues, il semble qu'il ne s'agisse sans doute pas d'une GAPDH.

Chez les levures, trois gènes ont été clonés et séquencés chez *Saccharomyces cerevisiae* (Holland et Holland, 1979a et 1979b ; Holland et al., 1983). Une autre séquence très peu différente a été définie très récemment chez *Zygosaccharomyces rouxii* (Imura et al., 1987).

Les séquences primaires des GAPDH animales définies sont celles du porc (Harris et Perham, 1968), du homard (Harris et Waters, 1976), du poulet (Dugaiczyck et al., 1983), du rat et de l'homme (Tso et al., 1985a) et de la drosophile (Tso et al., 1985b). La séquence de l'enzyme de nématode a été déterminée récemment (Yarbrough et al., 1987).

Enfin, parmi les enzymes d'origine végétale, la structure primaire de la GAPDH cytosolique de moutarde a été définie à partir de la séquence nucléotidique du cDNA par Martin et Cerff (1986). Il en est de même des enzymes cytosoliques et chloroplastiques de tabac (Shih et al., 1986) et, plus récemment, de celles du maïs (Brinkmann et al., 1987).

Toutes ces séquences possèdent entre 330 et 350 amino-acides. Leur structure primaire est hautement conservée. Les études comparatives entre les différentes séquences primaires peuvent donc servir de base pour établir une filiation entre les espèces. Elles permettent également de définir la filiation entre les différentes enzymes au sein d'une même espèce, ce qui est particulièrement intéressant dans le cas des enzymes chloroplastiques.

1-2. Structure tertiaire

Plusieurs structures tridimensionnelles de GAPDH glycolytiques ont été étudiées par l'approche cristallographique, dont celles des enzymes de muscle de homard (Moras et al., 1975 ; Murthy, 1980) et celle de *Bacillus stearothermophilus* (Biesecker et al., 1977 ; Leslie et Wonacott, 1983), cette dernière étant définie à haute résolution (Skarzynski et al., 1987 et 1988).

Ces travaux ont permis de montrer l'existence dans chaque monomère de deux domaines distincts (Fig. 4).

1-2-1. Le domaine de fixation du coenzyme

Il comprend les 148 premiers amino-acides du côté N-terminal. Ce domaine correspond au site de fixation du NAD⁺. Il se compose de six feuillets β plissés parallèles et de deux hélices α , suivant la nomenclature dérivée des travaux de Rossman et al. (1973). Une structure similaire a été observée au niveau des sites de fixation du coenzyme d'autres déshydrogénases, telles l'alcool déshydrogénase, la lactate déshydrogénase ou la malate déshydrogénase. Les études cristallographiques ont montré que l'hélice α constituant l'extrémité C-terminale de l'enzyme à partir du résidu 312 se repliait sur le domaine N-terminal et pouvait donc être considérée comme un élément supplémentaire appartenant au domaine de fixation du coenzyme.

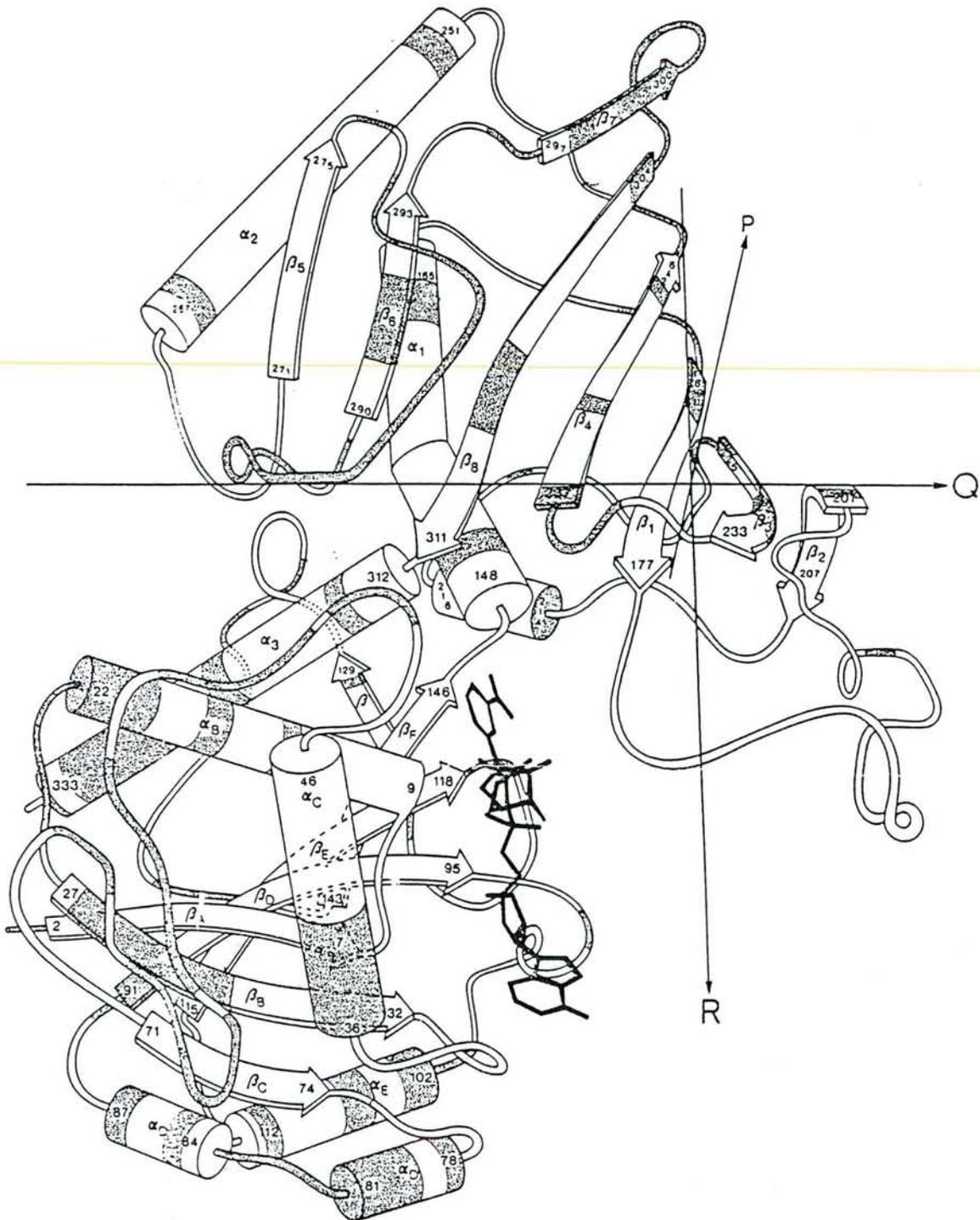


Figure 4 : Structure tridimensionnelle du monomère de la GAPDH de *B. stearothermophilus* d'après Walker et al. (1980)

1-2-2. Le domaine catalytique

Il comprend la partie comprise entre les résidus 149 et 311. Ce domaine se compose de huit feuillets β plissés antiparallèles et de trois hélices α approximativement parallèles (Rossman et al., 1973). Ce domaine comporte une région peu structurée, localisée entre les acides aminés 178 et 201, appelée "S-loop". Cette boucle constitue le "core enzyme" et semble impliquée dans les interactions entre les quatre sous-unités du tétramère (Walker et al., 1980b).

1-3. Symétrie moléculaire

La structure quaternaire des GAPDH a été définie dans un système de coordonnées à 3 axes P, Q, R orthogonaux.

Moras et al. (1975) ont postulé une asymétrie entre les sous-unités le long de l'axe Q de l'enzyme de homard qui se comporterait comme un dimère de dimère. Cette asymétrie se caractériserait par une conformation différente du NAD⁺.

Par contre, Biesecker et al. (1977) ont montré, à la résolution de 2,8 Å, une symétrie parfaite de 2.2.2 pour l'holoenzyme de *Bacillus stearothermophilus*, confirmée par les travaux plus récents réalisés à la résolution de 1,8 Å par Skarzynski et al. (1987) et ceux sur l'apoenzyme à la résolution de 2,5 Å (Leslie et Wonacott, 1983 et 1984 ; Skarzynski et Wonacott, 1988).

1-4. Relations structure-stabilité

Les GAPDH de *Bacillus stearothermophilus* et de *Thermus aquaticus*, bactéries thermophiles modérées, conservent leur activité enzymatique jusqu'à des températures de 70°C et 90°C respectivement, alors que dans les mêmes conditions, la GAPDH des organismes mésophiles est inactivée très rapidement.

La comparaison des structures primaires et tertiaires de GAPDH de mésophiles et de thermophiles a permis d'avancer un certain nombre d'hypothèses concernant les facteurs qui pourraient expliquer la thermostabilité du tétramère. Il s'agit de liaisons ioniques établies entre les résidus Arg 281 des deux sous-unités reliées par l'axe Q et les résidus Glu 201 des sous-unités reliées par les axes R et P. D'autres liaisons existent entre les résidus Arg 194 et Asp 293 des sous-unités reliées par l'axe P (Walker et al., 1980b).

Ces auteurs ont observé qu'en raison d'une conformation différente, ces dernières ne peuvent s'établir dans l'enzyme de homard et seraient spécifiques des enzymes de thermophiles. Skarzynski et Wonacott (1988) ont suggéré que l'Asn 180 puisse jouer un rôle dans la stabilité de la GAPDH à température élevée. Présent uniquement dans les enzymes de thermophiles, *B.stearotherophilus* et *T.aquaticus*, ce résidu pourrait contracter, par sa chaîne latérale, des liaisons hydrogène avec les atomes d'azote ou d'oxygène de la chaîne polypeptidique de la Leu 185 appartenant à la sous-unité séparée par l'axe R.

Selon Walker et al. (1980b), l'hydrophobie de certains résidus de la "S-loop" pourrait aussi jouer un rôle important dans les enzymes de thermophiles. Cette hypothèse semblait confortée par l'étude réalisée sur l'enzyme d'*Escherichia coli* qui présente davantage d'homologies avec les enzymes d'eucaryotes qu'avec les enzymes de bactéries thermophiles, notamment au niveau de la "S-loop" (Branlant et Branlant, 1985).

Cependant, lors d'études récentes effectuées sur la GAPDH chloroplastique de la moutarde, enzyme NADP⁺-dépendante et impliquée dans la photosynthèse, Martin et Cerff (1986) ont montré que la "S-loop" de cette GAPDH de mésophile présentait davantage d'homologies structurales avec la "S-loop" des GAPDH de bactéries thermophiles qu'avec celles des autres GAPDH séquencées jusqu'à ce jour, en particulier celle d'*E.coli*. Ceci semblait infirmer l'hypothèse de l'intervention de l'hydrophobie de la "S-loop" dans la thermostabilité.

La GAPDH d'*E.coli* étant la seule enzyme de bactérie mésophile connue lorsque nous avons commencé ce travail, il était nécessaire d'étudier d'autres enzymes bactériennes pour pouvoir disposer de davantage d'informations sur ce problème.

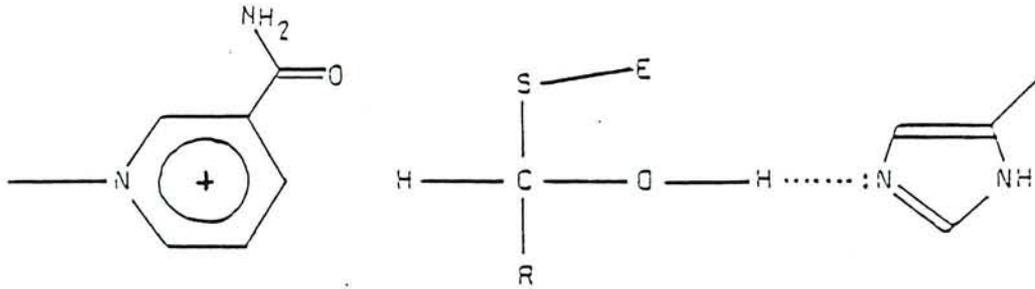
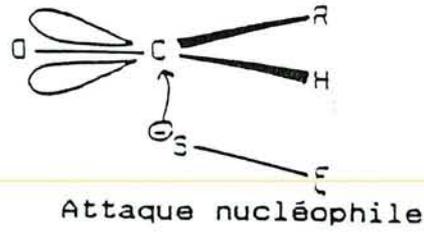
2. Mécanisme enzymatique de la GAPDH glycolytique

L'holoenzyme de la GAPDH glycolytique catalyse une réaction qui se compose d'une étape d'oxydo-réduction et d'une étape de phosphorylation.

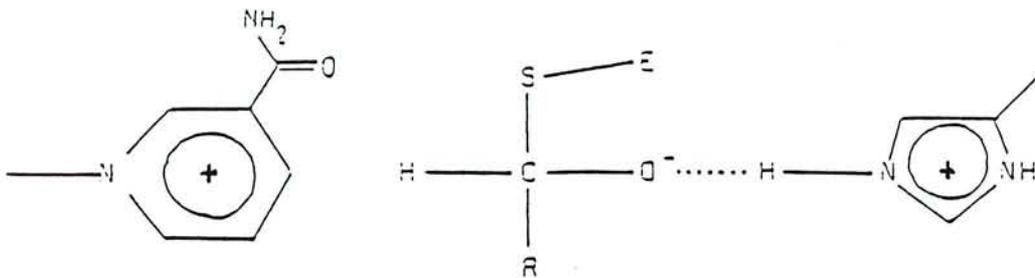
2-1. Formation de l'intermédiaire thiohémicétal (Fig. 5)

Après activation de la nucléophilie du thiol de la Cys 149, le carbone aldéhydique est attaqué par le thiolate, conduisant ainsi à la formation de l'intermédiaire thiohémicétal. Il se forme une liaison hydrogène entre le OH en C₁ de cet intermédiaire et l'azote N ε de l'His 176, ce qui favorise le processus d'oxydo-réduction.

Figure 5 : Formation de l'intermédiaire thiohémicétal



Formation de la liaison hydrogène entre l'hémithiocétal et l'His 176



Formation de la paire d'ions

2-2. Formation de l'acyl-enzyme (Fig. 6)

L'His 176 jouant le rôle de base, l'hydrogène en C₁ est transféré alors sur le C₄ du pyridinium. Le NAD⁺ est réduit en NADH qui se dissocie du complexe tertiaire tandis qu'une nouvelle molécule de NAD⁺ se forme. Ce processus d'échange de coenzyme NADH-NAD se matérialise par un fort changement conformationnel qui, d'après Wonacott (communication personnelle), conduirait le phosphate de l'acyl-enzyme du site de Pi au site Ps (voir paragraphe 2-4-2).

2-3. Formation de l'intermédiaire phosphorylé (Fig. 7)

Le phosphate inorganique réalise une attaque nucléophile sur l'intermédiaire thioacyle, facilitée par la présence de la liaison hydrogène entre le groupement C=O du thioester et l'imidazolium. Il se forme alors un intermédiaire tétraédrique conduisant à l'acide 1,3 diphosphoglycérique et à l'holoenzyme.

2-4. Les amino-acides impliqués dans ce mécanisme

Un certain nombre de résidus semblent jouer un grand rôle dans le mécanisme réactionnel de la GAPDH glycolytique. L'examen des structures primaires de GAPDH montre d'ailleurs que ces acides aminés sont fortement conservés au cours de l'évolution, ce qui confirme leur importance.

2-4-1. Ceux intervenant dans la catalyse enzymatique

La Cys 149 et l'His 176 jouent un rôle essentiel dans le processus de phosphorylation oxydative. Ils sont situés dans le site actif, à une distance d'une liaison hydrogène l'un de l'autre (Fig. 8), c'est-à-dire 3,96 Å dans l'apoenzyme et 3,63 Å dans l'holoenzyme.

- **La Cys 149** : intervient dans la formation des intermédiaires covalents thiohémicétal et thioacyle. Son rôle a été mis en évidence par modifications chimiques réalisées à l'aide d'agents, tel l'iodoacétate, qui conduisent à une perte totale de l'activité de la GAPDH (Segal et Boyer, 1953).

Figure 6 : Formation de l'acyl-enzyme

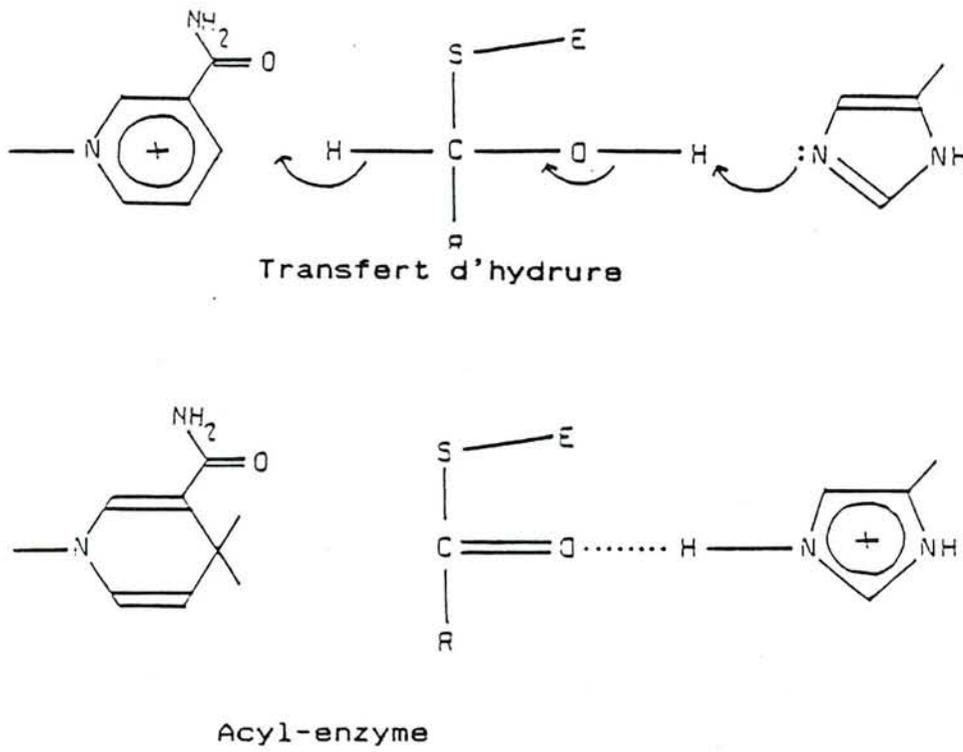
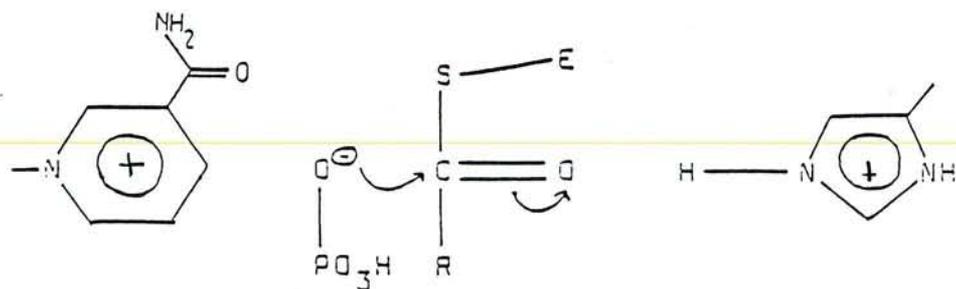
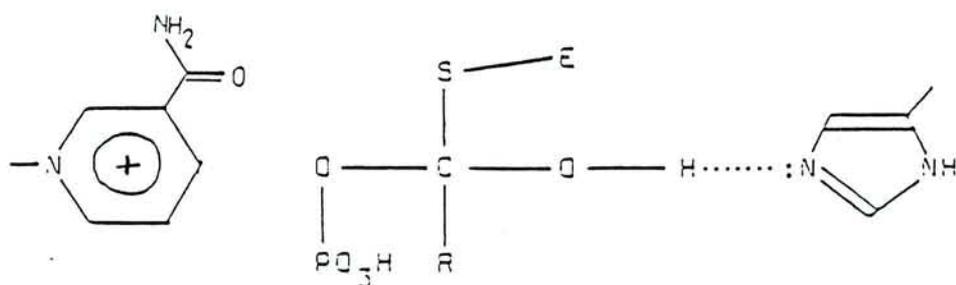


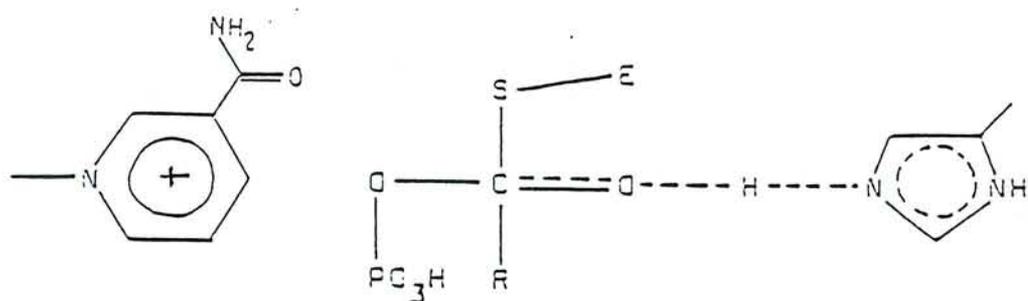
Figure 7 : Formation de l'intermédiaire phosphorylé



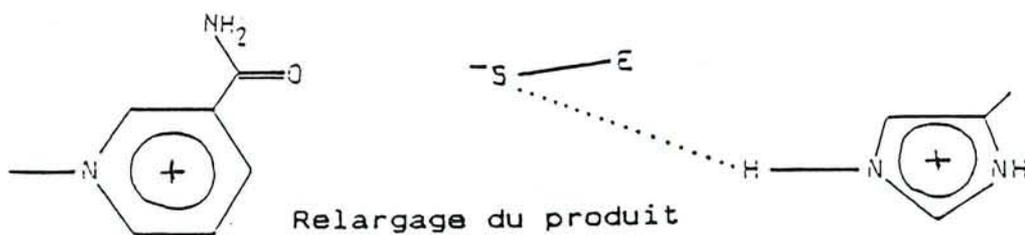
Attaque du phosphate sur l'acyl-enzyme



Intermédiaire phosphorylé



Etat de transition conduisant au relargage du produit



Relargage du produit

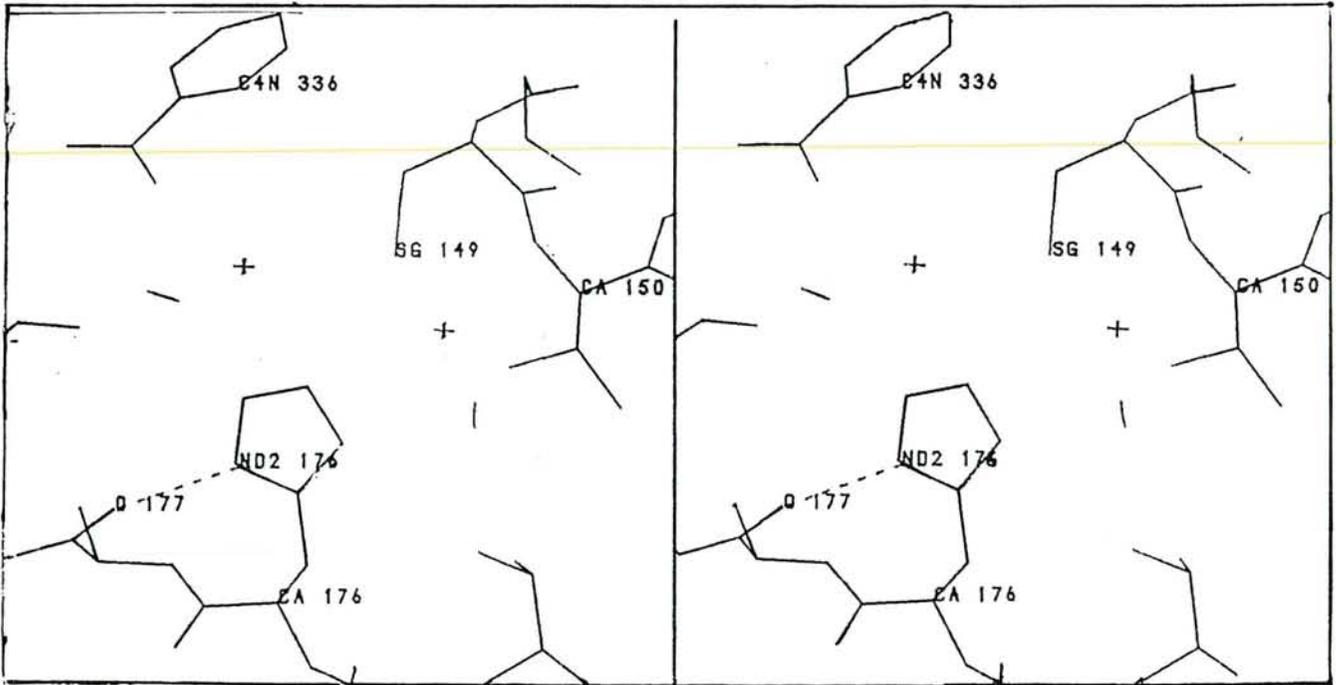


Figure 8 : Vue stéréo du site actif de la GAPDH de *B. stearothermophilus* d'après Wonacott A.J. (communication personnelle)

Ceci a été confirmé au laboratoire par substitution de la Cys 149 par mutagenèse dirigée. L'enzyme mutée possédant une glycine en position 149 est totalement inactive. L'activité catalytique est conservée dans l'enzyme dans laquelle une sérine a été substituée à la Cys 149, mais la constante catalytique n'équivaut qu'à 1/100^e de celle de l'enzyme sauvage (Soukri, 1988).

- **L'His 176** : intervient en exacerbant la réactivité chimique du groupement thiol de la Cys 149 en diminuant le pKa de celle-ci. Cette interaction, postulée par Polgar (1975), a été récemment confirmée par des études réalisées au laboratoire, en substituant par mutagenèse dirigée une Asn à l'His 176 (Soukri, 1988). Elle favorise aussi le transfert d'hydrure entre le C₁ du thiohémicétal vers le C₄ du pyridinium lors de l'étape conduisant à la formation de l'acyl-enzyme. Elle contribue également à la formation des intermédiaires et des états de transition correspondants, en particulier en augmentant le caractère électrophile du C₁ lors de l'attaque du phosphate inorganique sur l'acyl-enzyme (Harris et Waters, 1976).

2-4-2. Ceux impliqués dans la fixation du GAP et du phosphate inorganique

La D-spécificité du C₂ du substrat serait liée à une interaction de type liaison hydrogène contractée entre l'hydroxyle en C₂ et la Ser 148, le phosphate en C₃ favorisant cette spécificité par des interactions propres. Celui-ci interagirait avec les résidus Ser 148, Thr 150, Arg 195 et Thr 208 dans le site Pi.

Moras et al. (1975) ont mis en évidence l'existence de deux sites anioniques nommés Pi et Ps. D'après Skarzynski et al. (1987), au départ, le phosphate en C₃ du substrat serait fixé au site Pi (Fig. 9). L'étape d'isomérisation lors du changement NADH-NAD conduirait à un réarrangement conformationnel des sites Pi et Ps. Le phosphate du C₃ irait alors interagir avec les résidus Thr 179 et Arg 231 du site Ps. Ceci libèrerait le site Pi, permettant la fixation, puis l'attaque nucléophile, du phosphate inorganique sur l'intermédiaire thioacyle. Des études de mutagenèse dirigée sont actuellement entreprises au laboratoire pour confirmer cette hypothèse.

2-4-3. Ceux impliqués dans la fixation du NAD et ceux responsables de la spécificité de cofacteur

Le coenzyme se positionne dans son site par le jeu coordonné de plusieurs interactions avec les résidus constituant ce site (Fig. 10).

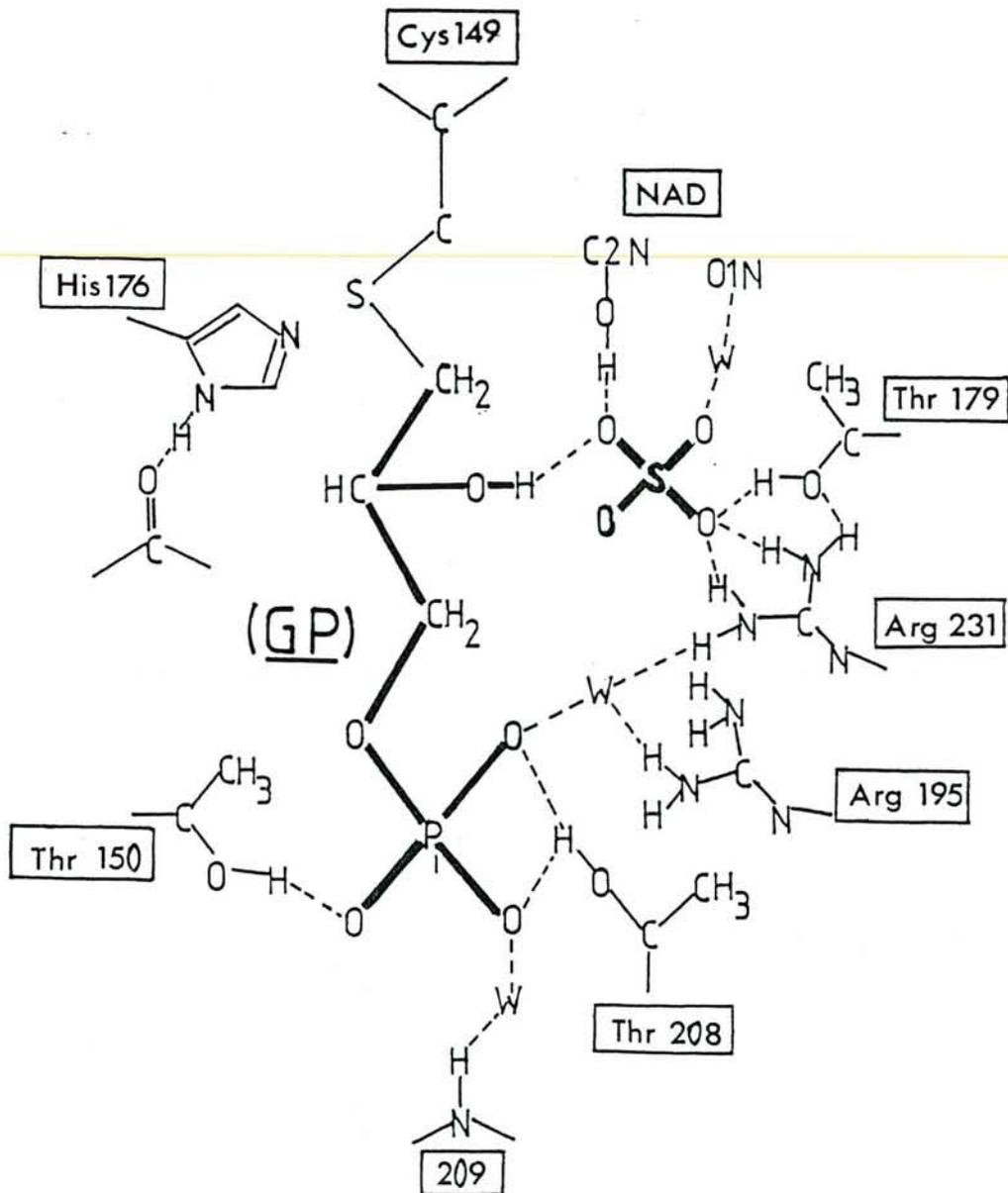


Figure 9 : Fixation du glycidol-3-phosphate dans le site actif de la forme holoenzyme de la GAPDH de *B. stearothermophilus* d'après Moody P.C.E. (résultats non publiés)

Les études cristallographiques ont permis de distinguer trois sous-sites différents engagés dans des interactions avec des parties précises de la molécule de NAD⁺ (Skarzynski et Wonacott, 1988).

Le premier de ces sous-sites est constitué par la chaîne latérale de l'Asn 313 qui fixe l'extrémité nicotinamide du NAD.

La chaîne latérale de la Ser 119 forme une liaison hydrogène avec le ribose du nicotinamide. Elle constitue le second sous-site.

Le troisième sous-site, le plus large, est constitué des Arg 10, Ile 11, la chaîne latérale de Asp 32 et Arg 77. Dans l'holoenzyme, ce sous-site est impliqué dans la fixation de la partie du coenzyme correspondant à la partie AMP. Il forme des liaisons hydrogène avec le groupement pyrophosphate par ses résidus Arg10 et Ile 11, avec le ribose par l'Asp 32 (résidu conservé aussi bien dans les GAPDH glycolytiques que chloroplastiques), avec l'adénine par l'Arg 77.

Dans les enzymes de thermophiles, l'Asn 180 est impliquée dans une liaison hydrogène supplémentaire avec le pyrophosphate du NAD.

Pour permettre le jeu complet de toutes ces interactions entre l'enzyme et la partie ADP de la molécule de NAD⁺, les résidus Asp 32 et Leu 33 subissent un changement conformationnel. Celui-ci permet à l'Asp 32 de former deux liaisons hydrogène avec le ribose, tandis que la chaîne latérale de la Leu 33 effectue une rotation et l'ensemble de ce résidu se déplace vers le noyau adénine du NAD avec lequel il établit des interactions hydrophobes. D'autre part, ces modifications permettent au noyau nicotinamide de contracter des interactions de type hydrophobe avec Ile 11, Tyr 317 et de type électrostatique avec Ser 119 et Asn 313.

Les interactions de la molécule de NAD avec les trois sous-sites décrits conduisent à des mouvements de ceux-ci, entraînant un changement conformationnel de la molécule entière. L'une des conséquences les plus importantes est le rapprochement des résidus Cys 149 et His 176 au niveau du site actif de la GAPDH (voir paragraphe 2-4-1 qui précède), ce qui augmente la réactivité de la Cys 149.

Il semble que la spécificité de l'enzyme glycolytique, qui accepte le NAD⁺ et non le NADP⁺ comme cofacteur, puisse résulter de plusieurs phénomènes simultanés. La présence d'un groupement phosphate sur le ribose en OH 2' pourrait conduire à une répulsion électrostatique due au groupement carboxyle de l'Asp 32. Ce n'est pas le cas chez les enzymes chloroplastiques, malgré la présence de ce résidu, en raison d'un environnement différent qui pourrait éviter cette répulsion.

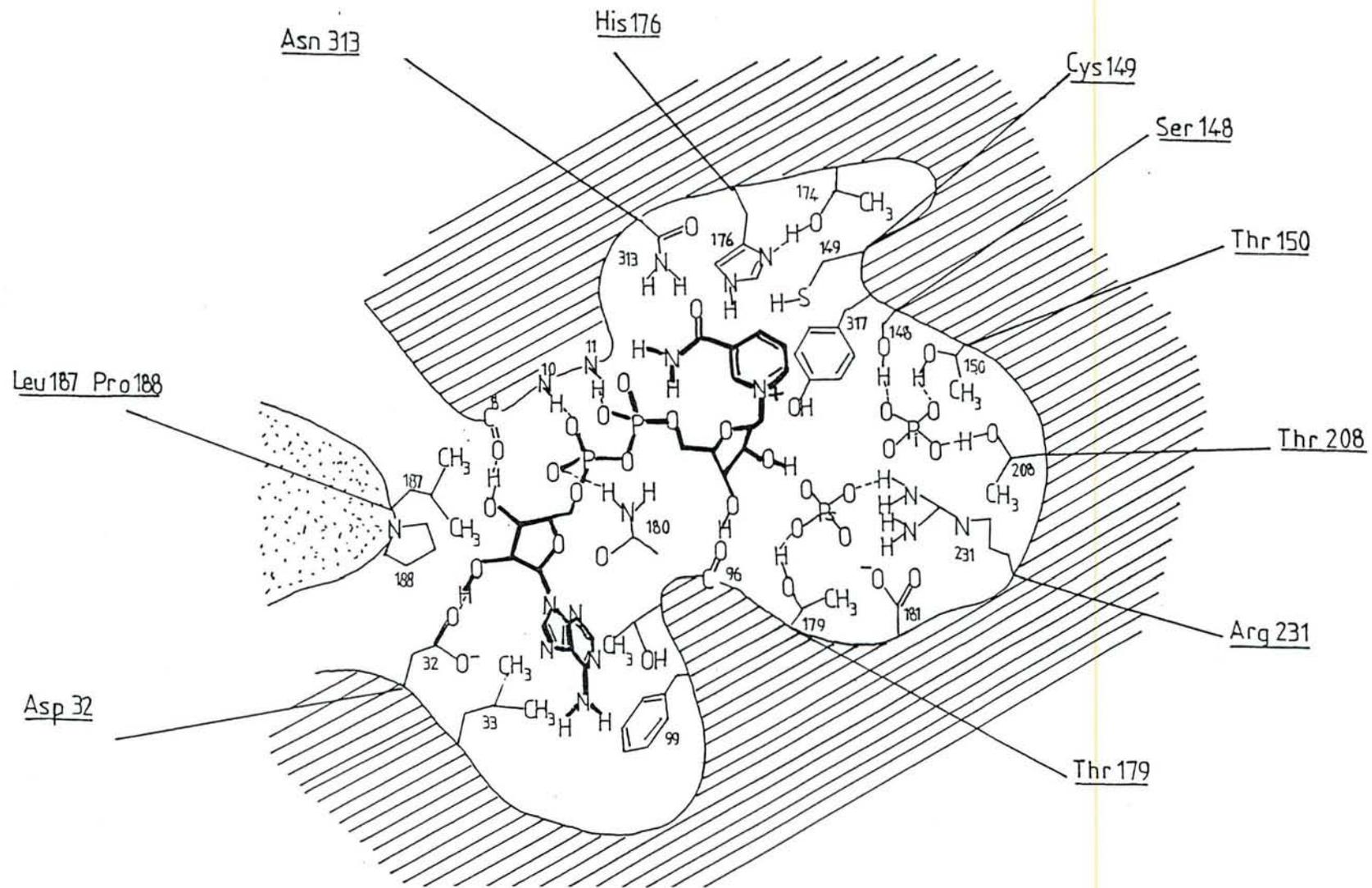


Figure 10 : Les sites anioniques et le site de fixation du NAD de la GAPDH de *B. stearothermophilus* d'après Biesecker et al. (1977)

De plus, la présence des résidus Leu 187 et Pro 188 de la sous-unité adjacente pourrait provoquer un encombrement stérique que subirait le groupement phosphate du NADP⁺, empêchant donc son positionnement correct dans le site de fixation du coenzyme. Cette hypothèse a été établie sur l'enzyme de *Bacillus stearothermophilus* à partir de comparaisons de structures primaires et de simulations graphiques sur ordinateur (Wonacott, communication personnelle). En effet, dans les enzymes chloroplastiques NADP⁺-dépendantes, la Pro 188 est absente. De plus, la Leu 187, présente dans les enzymes de thermophiles, est remplacée dans toutes les autres séquences par un résidu plus petit (Gly ou Ala).

Des travaux récents réalisés au laboratoire ont permis de changer la spécificité de cofacteur de l'enzyme glycolytique de *Bacillus stearothermophilus*. En effet, par mutagenèse dirigée, les résidus Leu187 et Pro 188 ont été remplacés par des amino-acides plus petits. L'enzyme mutée accepte le NADP⁺ comme cofacteur, confirmant l'implication de ces deux résidus dans la spécificité de coenzyme de la GAPDH glycolytique (Billard, 1988).

3. Etude de l'allostérie

A ce jour, différents travaux ont montré que les GAPDH étaient des enzymes allostériques possédant des propriétés de coopérativité caractérisées par une non équivalence des sites de fixation du coenzyme, comme c'est le cas chez la levure (Stallcup et Koshland, 1973a et 1973b).

Cette coopérativité est négative chez la levure (Conway et Koshland, 1968) ou le homard (Cook et Koshland, 1970). En effet, l'affinité pour le NAD⁺ diminue en fonction du taux d'occupation des sites. Cependant, dans certaines conditions de température, la fixation du NAD⁺ sur l'enzyme de levure peut se faire de façon coopérative positive (Kirschner et al., 1971), suggérant que la GAPDH existe en présence de NAD⁺ sous la forme active, thermodynamiquement plus stable.

Ces résultats montrent que la coopérativité entre sous-unités est partie intégrante de la réaction catalysée par la GAPDH, le monomère isolé ne présentant aucune efficacité enzymatique.

Ashmarina et al. (1985) ont montré que, chez la levure, la coopérativité observée entre les sous-unités de la GAPDH lors de la catalyse pouvait disparaître lors de la formation d'un complexe avec la phosphoglycérate kinase. Les auteurs ont suggéré que ce phénomène était à l'origine d'une régulation de l'activité enzymatique, un tel mécanisme moléculaire devant être effectif au sein de complexes constitués par des enzymes fonctionnellement reliées.

III - LES GENES DE GAPDH

Parmi les structures primaires de GAPDH connues aujourd'hui, plusieurs ont été déduites à partir des séquences nucléotiques des gènes dans le cas des procaryotes ou du cDNA dans le cas des enzymes eucaryotes.

Des études comparatives ont été effectuées sur les gènes eucaryotes, en particulier dans le cas des gènes d'enzymes de plantes.

1. Les gènes de GAPDH du règne animal

De nombreux travaux ont mis en évidence l'existence de plusieurs gènes actifs dans une même espèce, ce qui se traduit par la présence de plusieurs isoformes de GAPDH.

Chez *Trypanosoma brucei*, la présence de deux isoenzymes très différentes, l'une cytosolique, l'autre glycosomale, suggère l'existence de deux gènes actifs (Misset et al., 1987). Chez le nématode, au moins quatre gènes sont présents (Yarbrough et al., 1987). Deux gènes différents s'expriment chez la drosophile (Sullivan et al., 1985 ; Tso et al., 1985b).

Chez le poulet, les gènes de la GAPDH du muscle squelettique et de celle du muscle cardiaque ont été étudiés et comparés (Panabières et al., 1983) ; (Domdey et al., 1983). Ils présentent entre eux un seul nucléotide différent, se traduisant au niveau de la structure primaire par un changement de l'acide-amino 196, Asp dans le muscle squelettique et Gly dans le myocarde.

Arcari et al. (1984) ont eux déterminé la séquence de la GAPDH hépatique humaine et mis en évidence l'existence de plusieurs espèces de messagers, transcrits ou maturés différemment au cours du développement de l'individu. Cependant, aucune différence n'a été observée entre les structures primaires des enzymes du foie (Tso et al., 1985a) et du muscle squelettique (Hanauer et Mandel, 1984).

Les séquences nucléotidiques des gènes de GAPDH apportent un grand nombre d'éléments permettant d'analyser les mécanismes d'évolution de la protéine et du gène, en particulier l'évolution des introns. Ainsi, en comparant la séquence du DNA génomique avec la séquence du messenger de la GAPDH, Stone et al. (1985) ont pu déceler 11 introns chez le poulet. Parmi ceux-ci, certains sont localisés à la jonction de parties codant pour des domaines bien définis de l'enzyme.

En particulier, l'intron VI se situe entre les exons correspondant à la jonction entre le domaine de fixation du coenzyme et le domaine catalytique. L'intron 1 dans le gène de la GAPDH du nématode (Yarbrough et al., 1987) est localisé au même endroit. Ceci suggère l'existence de cet intron dans le gène ancestral. Il a été montré récemment l'existence de ce même intron dans le gène nucléaire de la GAPDH chloroplastique de maïs (Quigley et al., 1988).

Les signaux de transcription des gènes eucaryotiques de GAPDH sont des signaux classiques : en 5', une "CAAT" box, associée à une "TATA" box ; en 3', la séquence signal consensus de polyadénylation AATAAA. Les séquences nucléotidiques des gènes de GAPDH animales sont très conservées, puisque l'on constate 67 % d'homologie entre les gènes du nématode, du poulet ou de la drosophile. Ceci se retrouve d'ailleurs dans la forte homologie structurale constatée entre les enzymes du règne animal.

L'analyse du DNA génomique a, par ailleurs, permis de constater l'existence chez les vertébrés d'une grande multiplicité de séquences de DNA homologues au gène de la GAPDH (Hanauer et Mandel, 1984 ; Piechaczyk et al., 1984). Ces séquences très abondantes ne contiennent pas d'introns et ne s'expriment pas. Il s'agit de pseudogènes, estimés au nombre de 20 chez l'homme, mais existant à 200 copies au moins chez la souris et le rat. Ces séquences se caractérisent par une homologie très importante, environ 90 %, avec la séquence du cDNA obtenu à partir du mRNA de la GAPDH. Une hypothèse pouvant expliquer l'origine de ces pseudogènes pourrait être l'insertion de séquences de DNA obtenues à partir des mRNA par transcription réverse.

2. Les gènes de GAPDH du règne végétal

Différentes GAPDH ont été caractérisées chez les végétaux supérieurs : deux chez le maïs (Brinkmann et al., 1987) et la moutarde (Martin et Cerff, 1986) et trois chez le tabac (Shih et al., 1986). Ces enzymes sont soit cytosoliques, soit chloroplastiques. Quelle que soit leur localisation intracellulaire, toutes sont codées par des gènes nucléaires.

L'étude moléculaire de ces enzymes a apporté de nouveaux éléments de réponse au problème de l'origine des chloroplastes. En effet, deux théories s'affrontent à ce sujet. La théorie symbiotique propose que le chloroplaste provienne d'un procaryote symbiotique (Margulis, 1970) ; dans ce cas, le gène de la GAPDH chloroplastique aurait été transféré au noyau à partir du génome de cet ancêtre. La seconde théorie dite de compartimentation propose, elle, que le chloroplaste soit apparu à la suite d'une compartimentation de la cellule ancestrale (Raff et Mahler, 1972). Selon cette théorie, les gènes codant pour chacune des enzymes cytosolique et chloroplastique(s) résulteraient de la duplication d'un gène nucléaire ancestral unique de la GAPDH. De nombreuses études comparatives

réalisées sur les RNA ribosomiaux avaient déjà permis de renforcer la théorie de l'origine bactérienne des chloroplastes (Machatt et al., 1981).

Les travaux de Martin et Cerff (1986) et Brinkmann et al. (1987) ont montré que les enzymes chloroplastiques présentaient une homologie plus faible vis-à-vis des enzymes cytosoliques que des enzymes bactériennes thermophiles. De même, elles sont bien plus semblables à ces enzymes bactériennes qu'aux enzymes cytosoliques animales. Ces résultats renforcent donc la théorie endosymbiotique, les gènes nucléaires codant pour les fonctions exclusivement chloroplastiques ayant vraisemblablement pour ancêtres les gènes d'un symbiote (Weeden, 1981).

3. Les gènes de GAPDH de levures

Les travaux de Holland et al. (1983) et de Imura et al. (1987) ont permis de mettre en évidence l'existence de trois gènes exprimés chez *Saccharomyces cerevisiae* et deux chez *Zygosaccharomyces rouxii*. Ces gènes sont indépendants et ne constituent pas des formes alléliques différentes. Les régions codantes des gènes de *S.cerevisiae* présentent entre elles des homologies très importantes (de 90 à 95 %).

4. Les gènes de GAPDH bactériennes

Comme nous l'avons déjà signalé, au début de notre travail, le seul gène de GAPDH séquencé était celui d'*E.coli* (Branlant et Branlant, 1985), les autres structures d'enzymes bactériennes, de *Bacillus stearothermophilus* et *Thermus aquaticus*, ayant été déterminées directement à partir de la protéine.

Le gène de la GAPDH d'*E.coli* est précédé d'une séquence complémentaire à la séquence Shine-Dalgarno à six nucléotides en 5' du codon d'initiation. Le promoteur de ce gène présente une particularité. En effet, deux séquences promotrices potentielles ont été identifiées. Cette structure particulière du promoteur transcriptionnel peut être liée à la très forte expression du gène cloné. La structure de ce gène montrait que la GAPDH d'*E.coli* présentait une homologie de séquence plus importante avec les enzymes eucaryotes (64-69 %) qu'avec les enzymes procaryotes (49-57 %), l'homologie étant plus importante dans le domaine catalytique (78 %) que dans le domaine de fixation du cofacteur (55 %).

Martin et Cerff (1986) ont classé les GAPDH en deux branches, l'une eucaryotique et l'autre constituée des enzymes de *B.stearothermophilus*, de *T.aquaticus* et des enzymes chloroplastiques.

En effet, les enzymes chloroplastiques présentent une forte homologie avec les enzymes de thermophiles. Par contre, l'enzyme d'*E.coli* appartient clairement à la branche eucaryotique. Pour expliquer cela, les auteurs ont proposé que les rapports existant entre *E.coli* et les eucaryotes supérieurs auraient favorisé la possibilité d'un transfert de gène dans le sens eucaryote vers procaryote. Dans ce contexte, lorsque nous avons débuté notre travail, il était important d'isoler et de séquencer d'autres gènes de GAPDH de bactéries mésophiles afin de posséder de nouveaux éléments de comparaison au niveau bactérien.

D'autres équipes ont fait la même démarche et, récemment, deux autres séquences de gènes de GAPDH bactériennes ont été déterminées : celles de *Thermoproteus tenax* (Hensel et al., 1987) et *Zymomonas mobilis* (Conway et al., 1987).

IV - LES MICRO-ORGANISMES CHOISIS

Il était important de choisir un micro-organisme mésophile afin de savoir si *E.coli* constituait un cas particulier chez les mésophiles. Nous avons pour cela choisi de cloner et de séquencer le gène de la GAPDH de *Clostridium pasteurianum*. Ce choix était aussi intéressant par le fait qu'il s'agit d'un anaérobe strict sur lequel peu d'études moléculaires avaient été réalisées jusqu'alors.

Notre choix s'est porté d'autre part sur *Bacillus stearothermophilus*, bactérie thermophile dont la GAPDH avait fait l'objet de nombreuses études structurales et mécanistiques. La structure primaire de l'enzyme avait été déterminée directement sur la protéine et il était important de vérifier cette séquence à partir de celle du gène. De plus, peu d'éléments avaient été précisés sur l'expression des gènes de *B.stearothermophilus* et sur leurs promoteurs. Enfin, la structure tridimensionnelle de l'enzyme étant définie à haute résolution, il était intéressant de connaître la séquence nucléotidique pour pouvoir entreprendre une étude de la relation structure-fonction par mutagenèse dirigée.

1. Les signaux de régulation des gènes de bactéries Gram-positif

Quelques études ont porté sur les signaux de régulation chez les organismes Gram-positif en général. Ainsi, Graves et Rabinowitz (1986) ont-ils observé chez ces organismes des séquences promotrices analogues à la séquence -10, décrite par Pribnow (1975), et à la séquence -35 (Takanami et al., 1976 ; Seeburg et al., 1977). Les séquences consensus établies chez les bactéries Gram⁺ présentent toutefois des différences avec celles d'organismes Gram⁻ tel *E.coli*.

Elles sont par ailleurs accompagnées d'autres segments placés à proximité, telles la séquence riche en A et des séquences riches en A et T en amont de la séquence -35. Ces éléments seraient indispensables à une forte expression des gènes (Nishi et Itoh, 1986).

Les sites de terminaison de la transcription des gènes des bactéries Gram⁺ sont analogues à ceux des gènes d'*E.coli*, composés d'une séquence palindromique située en aval de la région codante, riche en G + C et suivie parfois d'une séquence polyT s'il s'agit d'un signal Rho-indépendant (Graves et Rabinowitz, 1986 ; Hinton et Freyer, 1986).

En ce qui concerne les signaux de traduction, les mRNA de la plupart des bactéries Gram⁺ présentent une séquence complémentaire à l'extrémité 3' fortement conservée du rRNA 16S. Cette séquence, nommée séquence SD (Shine et Dalgarno, 1975), constitue le site de fixation du ribosome sur le mRNA. Elle est distante de 5 à 14 nucléotides du codon d'initiation et influe sur l'efficacité de la traduction du messenger. La complémentarité de cette séquence avec le rRNA 16S est plus importante chez les bactéries Gram⁺ que chez les bactéries Gram⁻. De plus, une adénine située juste en 5' du codon d'initiation serait également impliquée dans la formation du complexe d'initiation de la traduction chez les bactéries Gram⁺ (Mc Laughlin et al., 1981).

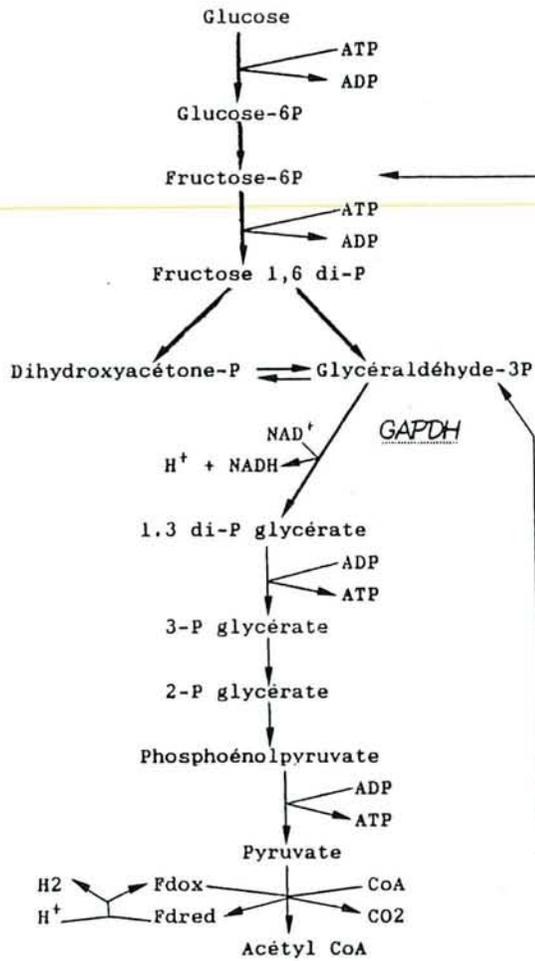
2. *Clostridium pasteurianum*

Le genre *Clostridium* regroupe l'ensemble des bactéries Gram-positif, anaérobies sporulantes, dont la spore est thermorésistante, centrale ou subterminale. *Clostridium pasteurianum* appartient au groupe des butyriques qui présentent un mécanisme fermentaire conduisant essentiellement à la production d'acides, ce qui explique l'intérêt industriel porté pour ces bactéries.

Cette espèce ne présente aucun pouvoir pathogène. C'est le fixateur d'azote le plus efficace des *Clostridia*. Il se présente sous la forme de bâtonnets droits et à bouts arrondis de 1,2 à 1,3 µm de large et 1,5 à 2 µm de long pour les éléments jeunes. Il se colore en jaune par l'iode et ne fermente pas le lactose.

La fermentation réalisée par ce micro-organisme emprunte quasi-exclusivement la voie du fructose-diphosphate (la voie glycolytique d'Embden-Meyerhoff-Parnas). Celle-ci produit deux molécules de pyruvate, d'ATP et de NADH par molécule d'hexose (Fig. 11).

Voie Glycolyse
EMBDEN-MEYERHOFF-PARNAS



Voie métabolique du xylose
(WARBURG-DICKENS)

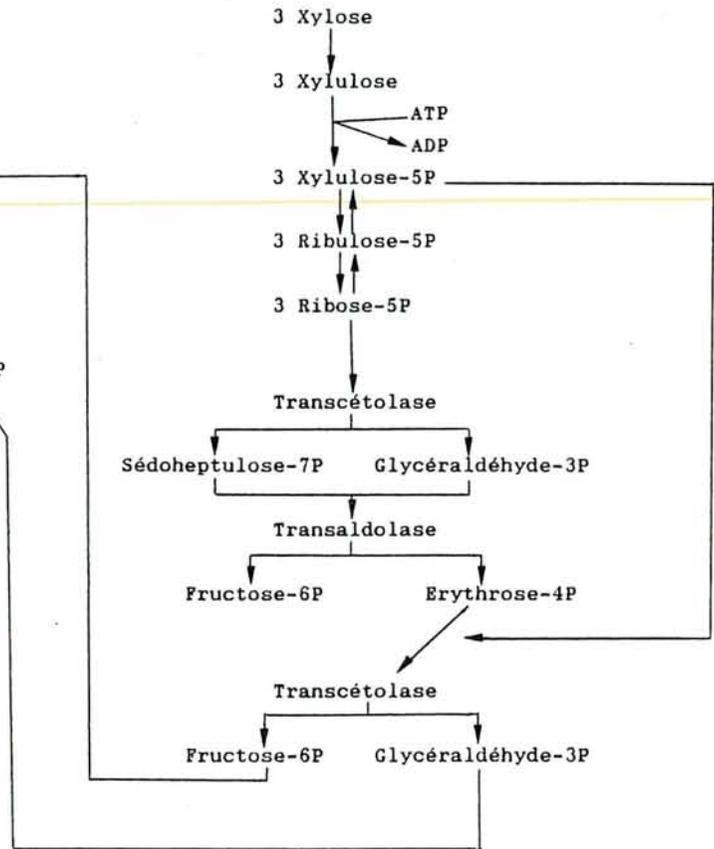


Figure 11 : Dégradation du glucose et du xylose chez *C.pasteurianum*

Le pyruvate est décarboxylé en acétyl-phosphate dans une réaction spécifique, la réaction phosphoroclastique (Fig. 12) dont le schéma global s'écrit ainsi :



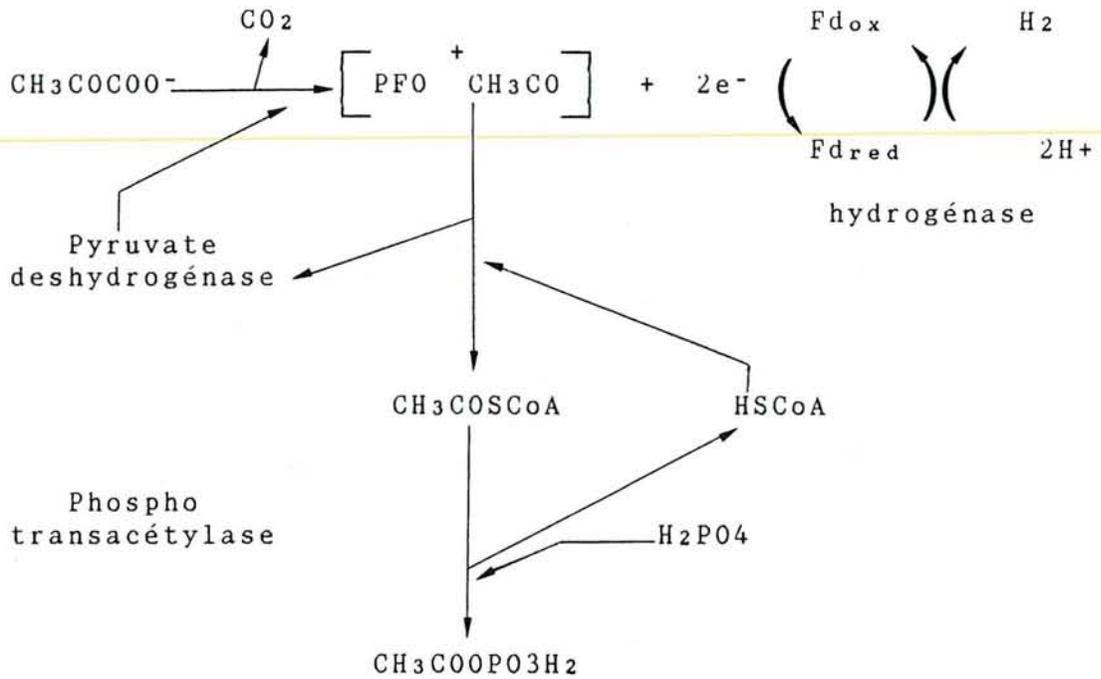
L'acétyl-phosphate ainsi produit à partir d'acétyl-CoA est transformé en acide acétique avec production d'une molécule d'ATP. Si la réaction phosphoroclastique est complète, deux acétyl-CoA se condensent en acétoacétyl-CoA qui sera transformé en butyryl-CoA avant de produire de l'acide butyrique et une molécule d'ATP (Fig. 13).

Quel que soit le produit de la fermentation, *C.pasteurianum* utilise toujours la même voie aboutissant au pyruvate, que ce soit à partir d'un hexose ou d'un pentose. La dégradation d'un pentose comme source carbonée fermentescible emprunte la voie de Warburg et Dickens qui produit du fructose-6-phosphate et du glycéraldéhyde-3-phosphate (Fig. 11). Ces derniers rejoignent alors la glycolyse, renforçant ainsi le rôle capital de la GAPDH chez cette bactérie.

Il faut préciser que quelques gènes clostridiens ont depuis été clonés, dont une partie sur la base de leur expression dans *E.coli*. Citons, chez *C.pasteurianum*, les gènes clonés et séquencés de la ferredoxine, transporteur d'électrons (Graves et al., 1985 ; Graves et Rabinowitz, 1986), de la galactokinase, enzyme catalysant une réaction au début de la voie métabolique du galactose (Daldal et Applebaum, 1985), des sous-unités de la nitrogénase, enzyme responsable de la fixation biologique d'azote (Chen et al., 1986 ; Wang et al., 1987) et de la protéine "mop" impliquée dans la fixation du molybdate (Hinton et Freyer, 1986).

3. *Bacillus stearothermophilus*

Micro-organisme aérobic Gram positif, *Bacillus stearothermophilus* présente la particularité de pouvoir se développer jusqu'à une température de 65°C. Il offre par ailleurs une tolérance limitée aux acides.



PFO : Pyruvate ferredoxine oxydo-réductase

Figure 12 : Réaction phosphoroclastique chez *C.pasteurianum*
(MORTENSON et al., 1963)

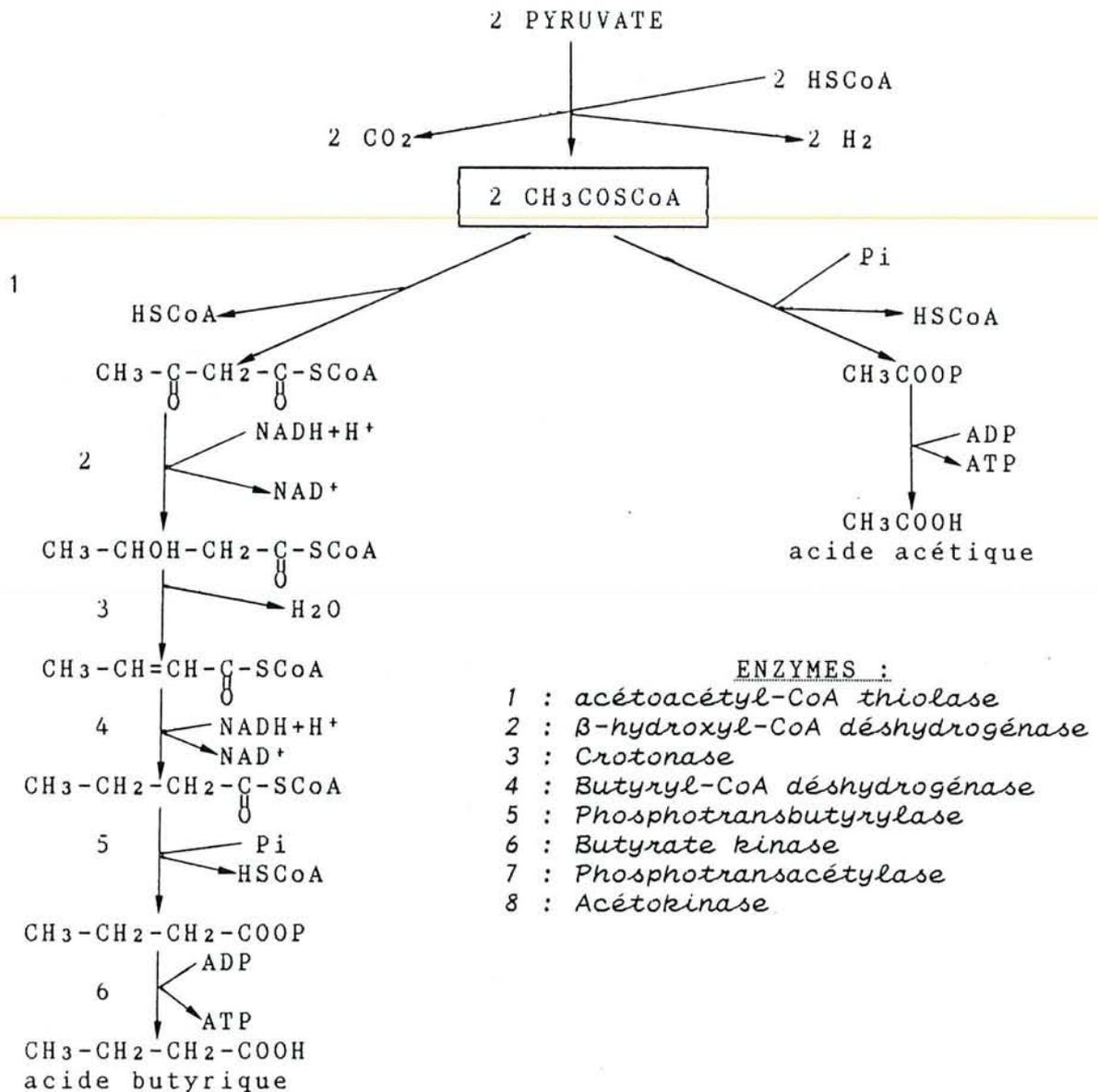


Figure 13 : Principales voies de dégradation de l'acétyl-CoA chez *C.pasteurianum*

Quelques gènes de *B.stearothermophilus* ont été clonés dans *E.coli* et séquencés. Il s'agit de ceux codant pour la tyrosyl-tRNA synthétase (Barker, 1982 ; Waye et Winter, 1986), la tryptophanyl-tRNA synthétase (Barstow et al., 1986a), la valyl-tRNA synthétase (Borgford et al., 1987), une protéase neutre (Takagi et al., 1985), l' α -amylase (Ihara et al., 1985), la L-lactate déshydrogénase (Barstow et al., 1986b) et le facteur d'initiation IF2 (Brombach et al., 1986). Ces gènes clonés ont été exprimés dans *E.coli* sous le contrôle de leurs promoteurs et malgré une utilisation des codons peu favorable dans *E.coli*. Peu d'études ont été réalisées sur les signaux de régulation propres à ces gènes. Lors de travaux menés sur les gènes de tyrosyl-tRNA synthétase et de tryptophanyl-tRNA synthétase, une séquence potentielle de terminaison de transcription Rho-indépendante a été identifiée à environ 50-60 nucléotides en 5' du codon d'initiation. Les auteurs pensent que cette structure pourrait être impliquée dans la régulation de la transcription de ces gènes (Waye et Winter, 1986 ; Barstow et al., 1986a).

MATERIELS ET METHODES

SOMMAIRE

I - MICRO-ORGANISMES ET MILIEUX UTILISES

1. <i>Clostridium pasteurianum</i>	page 48
1-1. Souche.....	page 48
1-2. Milieux et conditions de culture.....	page 48
2. <i>Bacillus stearothermophilus</i>	page 50
3. <i>Escherichia coli</i>	page 50
3-1. Souches.....	page 50
3-2. Milieux et conditions de culture.....	page 51
3-2-1. Milieu minimum M9.....	page 51
3-2-2. Milieu complet LB.....	page 52
3-2-3. Milieu minimum M63.....	page 52
3-2-4. Milieux solides et sélectifs.....	page 53

II - PLASMIDES ET PHAGES

1. Plasmides.....	page 53
1-1. pBR322.....	page 53
1-2. pUC18 et pUC19.....	page 55
1-3. Bluescript+ et Bluescript.....	page 55
2. Phages.....	page 55

III - TRANSFORMATION DES SOUCHES d'*E.coli*

1. Préparation de cellules compétentes.....	page 59
1-1. Souches BJ5183, HB101 et JM101.....	page 59
1-2. Souches DF221.....	page 59
2. Transformation de cellules compétentes.....	page 60
2-1. Par les plasmides.....	page 60
2-1-1. Souches BJ5183, HB101 et JM101.....	page 60
2-1-2. Souche DF221.....	page 60
2-2. Par les phages.....	page 61

IV - PREPARATION DE DNA

1. Préparation du DNA génomique de <i>C.pasteurianum</i>	page 61
2. Préparation du DNA génomique de <i>B.stearothermophilus</i>	page 62
3. Préparation de DNA plasmidique.....	page 63
3-1. Technique du lysat clair.....	page 63
3-2. Technique de lyse alcaline.....	page 63
4. Préparation du DNA du phage M13.....	page 64
4-1. Rappels sur la physiologie du phage M13.....	page 64
4-2. Forme simple brin.....	page 64
4-3. Forme double brin.....	page 66
5. Purification du DNA.....	page 66
5-1. Extractions au phénol et au chloroforme.....	page 66
5-2. Précipitation à l'éthanol.....	page 66

V - DIGESTION, DEPHOSPHORYLATION, SEPARATION ET LIGATION DE DNA

1. Digestion par les enzymes de restriction.....	page 67
2. Déphosphorylation du DNA.....	page 68
3. Séparation électrophorétique des fragments générés.....	page 68
3-1. En gel d'agarose.....	page 68
3-2. En gel de polyacrylamide.....	page 69
4. Electroélution des fragments séparés.....	page 70
5. Ligation.....	page 70

VI - PREPARATION DE DEOXY-OLIGONUCLEOTIDES DE SYNTHÈSE

1. Synthèse chimique.....	page 71
1-1. Principes généraux.....	page 71
1-2. Eléments de base.....	page 71
1-2-1. Le groupement hydroxyle en 5' du désoxyribose.....	page 72
1-2-2. Les bases.....	page 72
1-2-3. Le groupement phosphate en 3' du désoxyribose.....	page 72
1-3. La synthèse.....	page 72
1-3-1. Le support solide.....	page 72
1-3-2. Les réactions.....	page 73
2. Déprotection et purification de l'oligonucléotide.....	page 73
2-1. Solubilisation et déprotection.....	page 75
2-2. Purification par H.P.L.C.....	page 75
2-3. Détritylation de l'extrémité hydroxyle 5'.....	page 76

VII - SEQUENCAGE DE DNA

1. Technique sur le phage M13.....	page 76
1-1. Principes.....	page 76
1-2. Mise en évidence de clones recombinants du phage M13.....	page 78
1-3. Etapes.....	page 78
1-3-1. Solutions nécessaires au séquençage.....	page 78
1-3-2. Hybridation de l'amorce sur le DNA sb.....	page 80
1-3-3. Elongation.....	page 81
1-3-4. Chasse et arrêt de la réaction.....	page 81
2. Adaptation de la technique de Sanger au séquençage sur plasmide.....	page 81
3. Séparation électrophorétique.....	page 82
4. Traitement du gel et autoradiographie.....	page 83

VIII - MUTAGÈNESE DIRIGÉE

1. Principe de la technique.....	page 84
2. Etapes.....	page 84
2-1. Formation de l'hétéroduplex.....	page 84
2-2. Phosphorylation de l'oligonucléotide.....	page 86
2-3. Hybridation de l'oligonucléotide.....	page 87
2-4. Elongation et ligation.....	page 87
2-5. Transformation.....	page 87
3. Sélection des mutants.....	page 88

IX - HYBRIDATION DE DNA PAR LA METHODE DE SOUTHERN

1.Principe.....	page 88
2.Marquage d'une sonde de DNA par "random priming".....	page 88
2-1.Principe.....	page 88
2-2.Etapes.....	page 89
2-2-1. Dénaturation.....	page 89
2-2-2. Elongation.....	page 89
2-3.Purification de la sonde marquée.....	page 89
3.Transfert de DNA.....	page 90
3-1.Migration électrophorétique.....	page 90
3-2.Traitement du gel.....	page 90
3-3.Transfert sur membrane de nylon.....	page 90
3-4.Traitement de la membrane.....	page 90
4.Hybridation de la sonde.....	page 91
4-1.Préhybridation.....	page 91
4-2.Hybridation.....	page 91
4-3.Lavages.....	page 91
4-4.Révélation par autoradiographie.....	page 92

X - TESTS D'EXPRESSION DU GENE DE LA GAPDH

1.Préparation des échantillons.....	page 92
1-1.Culture cellulaire.....	page 92
1-2.Préparation de l'extrait brut.....	page 92
2.Estimation de la quantité d'enzyme.....	page 93
2-1.Composition du gel.....	page 93
2-2.Electrophorèse.....	page 94
2-3.Révélation et étude des profils.....	page 94
3.Mesure de l'activité enzymatique de la GAPDH.....	page 94
3-1.Solutions nécessaires.....	page 95
3-1-1. Tampon Ferdinand.....	page 95
3-1-2. Préparation du GAP.....	page 95
3-2.Test enzymatique.....	page 95

I - MICRO-ORGANISMES ET MILIEUX UTILISES

1. *Clostridium pasteurianum*

1-1. Souche

Nous avons utilisé la souche W5 (ATCC 6013) de *Clostridium pasteurianum*, bactérie anaérobie stricte, Gram-positif, isolée par Winogradsky en 1895.

1-2. Milieux et conditions de culture

Nous disposons de spores de *C.pasteurianum* conservées dans un milieu riche, le milieu RCM (Reinforced Clostridial Medium), commercialisé par Biolyon et dont voici la composition :

Extrait de levure.....	3	g
"Lab-lemco".....	10	g
Peptone.....	10	g
Amidon soluble.....	1	g
Dextrose.....	5	g
Cystéine hydrochloride.....	0,5	g
NaCl.....	5	g
CH ₃ COONa.....	3	g
Agar.....	0,5	g

pH..... 6,8

(quantités définies pour un litre d'eau distillée).

La germination des spores est réalisée dans ce même milieu après un choc thermique de 20 mn à 80°C suivi d'un refroidissement brutal dans la glace fondante. Les spores sont alors incubées 24 h à 34°C.

Nous avons cultivé *C.pasteurianum* sur deux milieux différents, soit synthétique, soit complexe.

Le milieu synthétique a été décrit par Lovenberg et Sobel (1965).

MgSO ₄ , 7 H ₂ O.....	175	mg
NaCl.....	100	mg
(NH ₄) ₆ Mo ₇ O ₂₄ , 4 H ₂ O.....	7,4	mg
KH ₂ PO ₄	1,4	g
K ₂ HPO ₄	15,6	g
FeSO ₄ , 7 H ₂ O (dans H ₂ SO ₄ 0,1 M).....	18	mg
SO ₄ (NH ₄) ₂	2	g
Biotine.....	0,01	mg
Saccharose.....	20	g

(quantités définies pour un litre d'eau distillée.)

Le milieu complexe a été optimisé par Monot et al. (1982).

MgSO ₄ , 7 H ₂ O.....	0,2	g
Extrait de levure.....	5	g
CaCO ₃	3	g
KH ₂ PO ₄	0,5	g
K ₂ HPO ₄	0,5	g
FeSO ₄ , 7 H ₂ O (dans H ₂ SO ₄ 0,1 M).....	10	mg
SO ₄ (NH ₄) ₂	2	g
Biotine.....	0,01	mg
Acide para-aminobenzoïque.....	1	mg
Glucose.....	60	g

(quantités définies pour un litre d'eau distillée.)

Quel que soit le milieu, la croissance de *C.pasteurianum* est toujours réalisée à 34°C en anaérobiose. Les cellules, en culture dans des flacons de 10 litres pleins, ne sont pas agitées et se développent au fond du récipient où règnent des conditions anaérobies.

Nous avons préparé le DNA génomique à partir de cellules de douze heures de culture sur milieu synthétique, c'est-à-dire en début de phase exponentielle de croissance.

2. Bacillus stearothermophilus

Nous avons utilisé la souche NCA 1503 de *Bacillus stearothermophilus* (Atkinson et al., 1979), bactérie aérobie, Gram-positif.

Nous disposons de cellules congelées de *B.stearothermophilus* à partir desquelles nous avons préparé le DNA génomique.

3. Escherichia coli

3-1. Souches

E.coli **JM101** : ($\Delta\lambda$ lac, pro), supE, thi-, strA, endA, sbcB15, hsdR4, F', traD36, proAB, LacIq, Z Δ M15, RecA-.

Cette souche possède sur son épisode le facteur F' nécessaire à l'infection par le phage M13. Cette souche est conservée sur milieu minimum glucosé pour éviter la perte du facteur F' (Maniatis et al., 1982).

E.coli **BJ5183** : F-, RecBC-, sbcB-, endA-, met-, thi-, bio-, gal-, res-, hsdR, strR.

Cette souche présente une excellente aptitude à être transformée par les plasmides. C'est dans cette souche que nous avons choisi de construire la banque génomique de *C.pasteurianum*.

E.coli **HB101** : F-, thi-, leu-, pro-, hsdM-, RecA-, supE (suII).

Cette souche possède la particularité de bien se transformer par les plasmides et de présenter le phénotype RecA- qui minimise les processus de recombinaison.

E.coli **HB2154** : ara, (Δ lac, pro), thi-, F', proA+B-, LacIq, Z Δ M15, mutL :: Tn10.

Cette souche est réceptrice du phage M13. Elle est mutée dans un gène majeur du système de réparation et ne possède pas de tRNA suppresseurs (Carter et al., 1985). Elle est utilisée pour la mutagénèse dirigée, étant incapable de corriger le mauvais appariement de base localisé au niveau de la mutation introduite.

E.coli **HB2151** : ara, (Δ lac, pro), thi⁻, F', proA⁺B⁻, LacI^q, Z Δ M15, mutL :: Tn10.

Cette souche possède les mêmes caractéristiques que la précédente, mais dispose, elle, d'un système de réparation fonctionnel qui lui permet de corriger les mutations spontanées comme les autres souches (Carter et al., 1985). Elle est utilisée comme tapis cellulaire lors des transformations de mutagénèse dirigée.

E.coli **DF221** :

Cette souche est déficiente en GAPDH, en raison d'une mutation ambre présente au niveau du gène structural de cette enzyme. Elle est utilisée pour le clonage par complémentarité et les études d'expression du gène de la GAPDH.

3-2. Milieux et conditions de culture

3-2-1. Milieu minimum M9

Ce milieu est utilisé pour la conservation des souches. Pour un litre de milieu :

Na ₂ HPO ₄ , H ₂ O.....	6,9	g
KH ₂ PO ₄	3,0	g
NaCl.....	0,5	g
NH ₄ Cl.....	1,0	g

Le mélange est autoclavé 30 mn à 1 bar.

Puis sont ajoutés stérilement :

MgSO ₄ 1M.....	2,0	ml
CaCl ₂ 1M.....	0,1	ml
Glucose 20 % (p/v).....	20,0	ml
Thiamine filtrée stérilement (20 mg/ml)	0,5	ml

Le milieu M9 gélosé est obtenu en ajoutant 20 g de bacto-agar avant la stérilisation.

3-2-2. Milieu complet LB

Pour un litre de milieu :

Peptone tryptique (Biofit).....	5	g
Peptone pancréatique (Biofit).....	5	g
Extrait de levure (Biofit).....	5	g
NaCl.....	5	g

Le pH est ajusté à 7,5 avec NaOH 1 N. Le milieu est alors stérilisé 55 mn à 1 bar.

3-2-3. Milieu minimum M63

Ce milieu est utilisé pour les cultures de la souche *E.coli* DF221.

Composition du milieu 2x concentré :

KH ₂ PO ₄	0,2	M
(NH ₄) ₂ SO ₄	0,03	M
MgSO ₄ , 7 H ₂ O.....	1,6	mM
FeSO ₄ , 7 H ₂ O.....	4,0	M

Le pH de cette solution est ajusté à 7,0 avec KOH 10 N. Elle est alors stérilisée 45 mn à 1 bar.

Cette solution M63 2x sert de base à la préparation de milieux sélectifs tels que le milieu M63 glucose dont la composition pour 100 ml est :

M63 2x.....	50	ml
Trp/Met 0,125 % (p/v).....	2	ml
Thiamine 0,2 % (p/v).....	50	μl
Casaminoacides 2,5 % (p/v).....	4	ml
Glucose 20 % (p/v).....	2	ml

Pour préparer le même volume de M63 succinate-glycérol, la solution de glucose est remplacée par une solution contenant 20 % (p/v) de succinate et 4,5 % (v/v) de glycérol.

3-2-4. Milieux solides et sélectifs

Les milieux solides sont obtenus en ajoutant 15 g de bacto-agar par litre avant stérilisation.

Pour la gélose molle, 6 g de bacto-agar sont ajoutés par litre de milieu LB.

Les milieux sélectifs sont préparés en ajoutant, après stérilisation, l'un des antibiotiques dans les proportions suivantes :

- Ampicilline : 100 mg/l de milieu
- Tétracycline : 15 mg/l de milieu.

Les cultures d'*E.coli* sont réalisées en aérobose sous agitation à 37°C. La croissance est suivie par dosage spectrophotométrique à 700 nm, sachant que 1 DO₇₀₀ = 10⁹ bactéries/ml.

Les ensemencements sont effectués au 1/100^e à partir de précultures saturées. Les cultures de la souche d'*E.coli* DF221, réalisées en milieu M63 succinate-glycérol (en milieu M63 glucose si elle contient un gène actif de GAPDH), sont ensemencées au 1/50^e à partir d'une préculture saturée.

II - PLASMIDES ET PHAGES

1. Plasmides

1-1. pBR322 (Bolivar et al., 1977 ; Bolivar et Backmann, 1979)

Le plasmide pBR322 renferme 4,3 kb. Il possède deux gènes conférant à la cellule-hôte une résistance à l'ampicilline et à la tétracycline (Fig. 14).

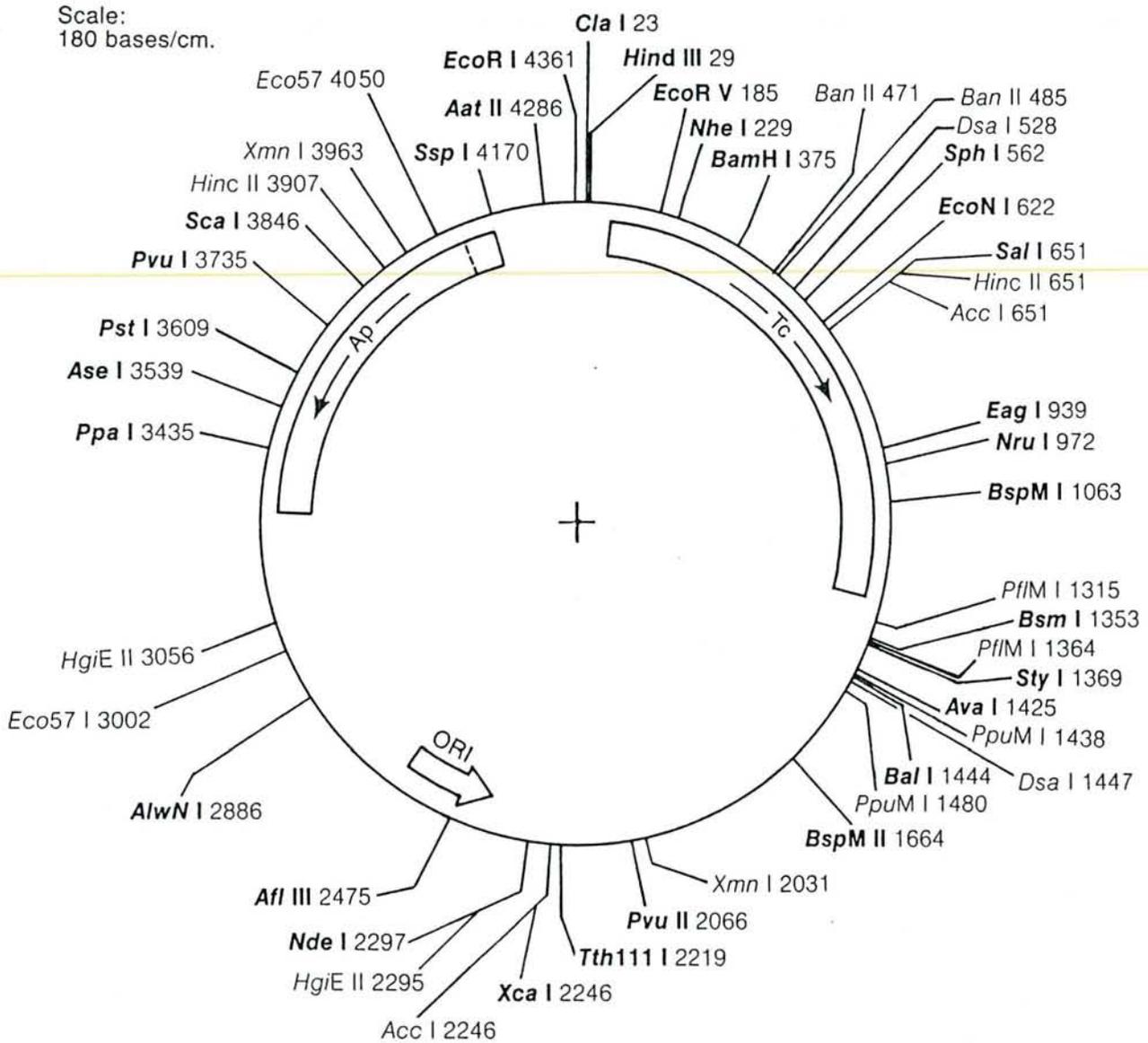


Figure 14 : Cartographie du plasmide pBR322

1-2. pUC18 et pUC19 (Vieira et Messing, 1982)

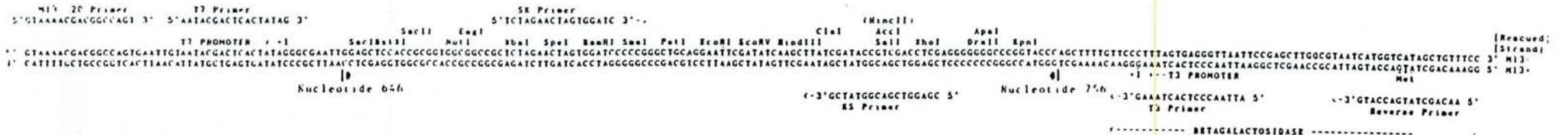
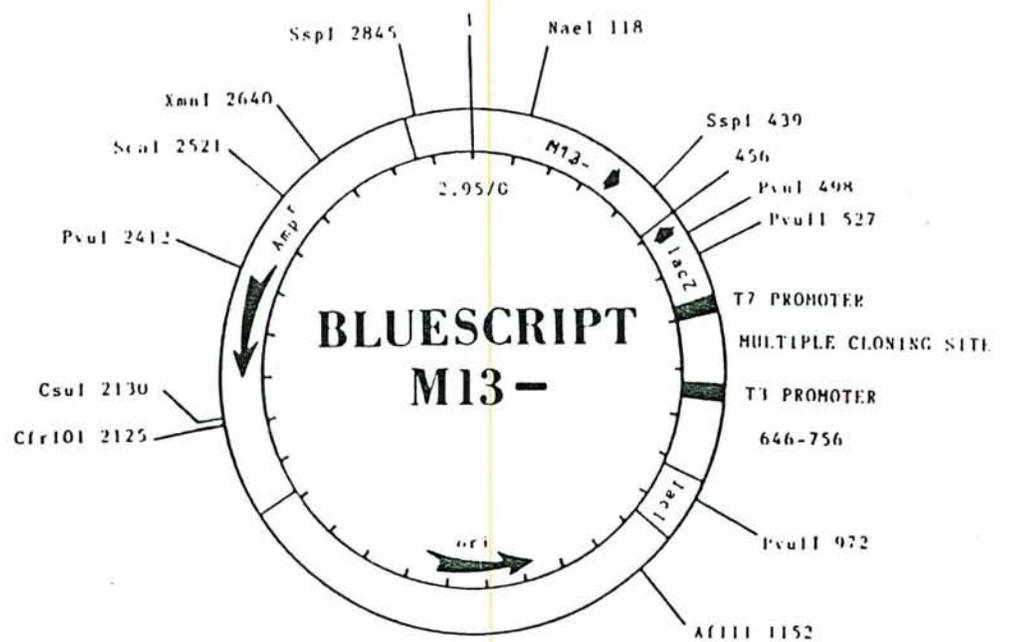
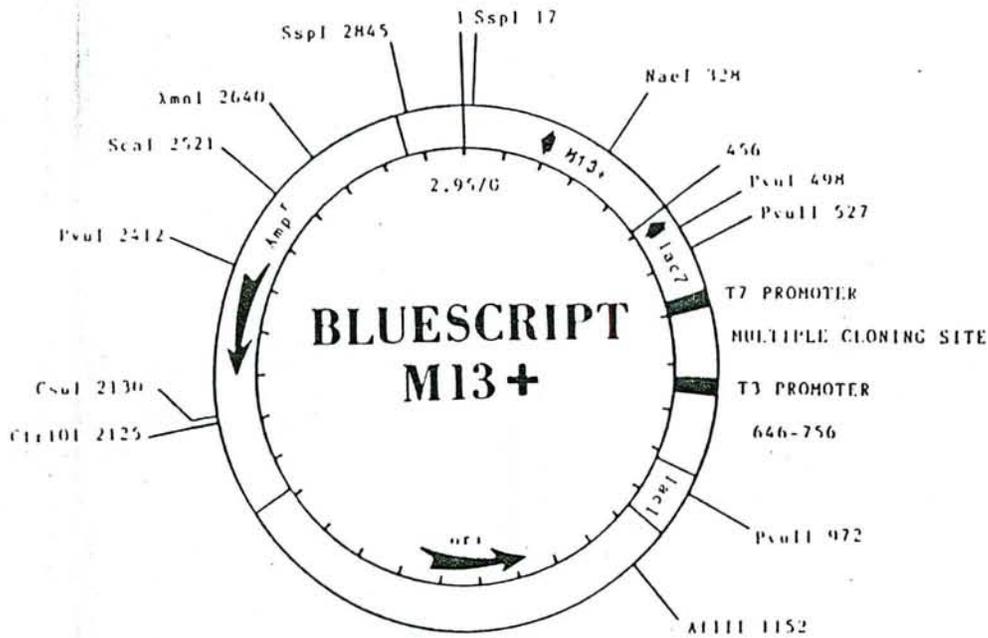
Ces plasmides renferment 2,7 kb. Ils dérivent de pBR322 duquel ils ont conservé l'origine de réplication et le gène de résistance à l'ampicilline (Fig. 15). Ils contiennent une cassette composée d'un fragment du gène de la β -galactosidase d'*E.coli* permettant l' α -complémentation de souches ayant la caractéristique génotypique Δ lac. Dans ce fragment du gène de la β -galactosidase a été insérée une séquence "polylinker" renfermant un grand nombre de sites uniques de coupure. Les plasmides pUC18 et pUC19 diffèrent uniquement par l'orientation de leur séquence "polylinker".

1-3. Bluescript+ et Bluescript- (Stratagene Inc., CA)

Ces plasmides de 2,96 kb possèdent de nombreux points communs avec les pUC. Ils renferment le gène de résistance à l'ampicilline et l'origine de réplication de pBR322, et le même type de cassette, la séquence "polylinker" possédant un plus grand nombre de sites (Fig. 16). Les plasmides Bluescript contiennent également un tronçon de DNA du phage M13 situé en aval du gène lac Z. Ce fragment a une orientation opposée dans Bluescript+ et Bluescript-. Il permet d'obtenir ces plasmides sous la forme simple brin par co-infection avec un autre phage.

2. Phages

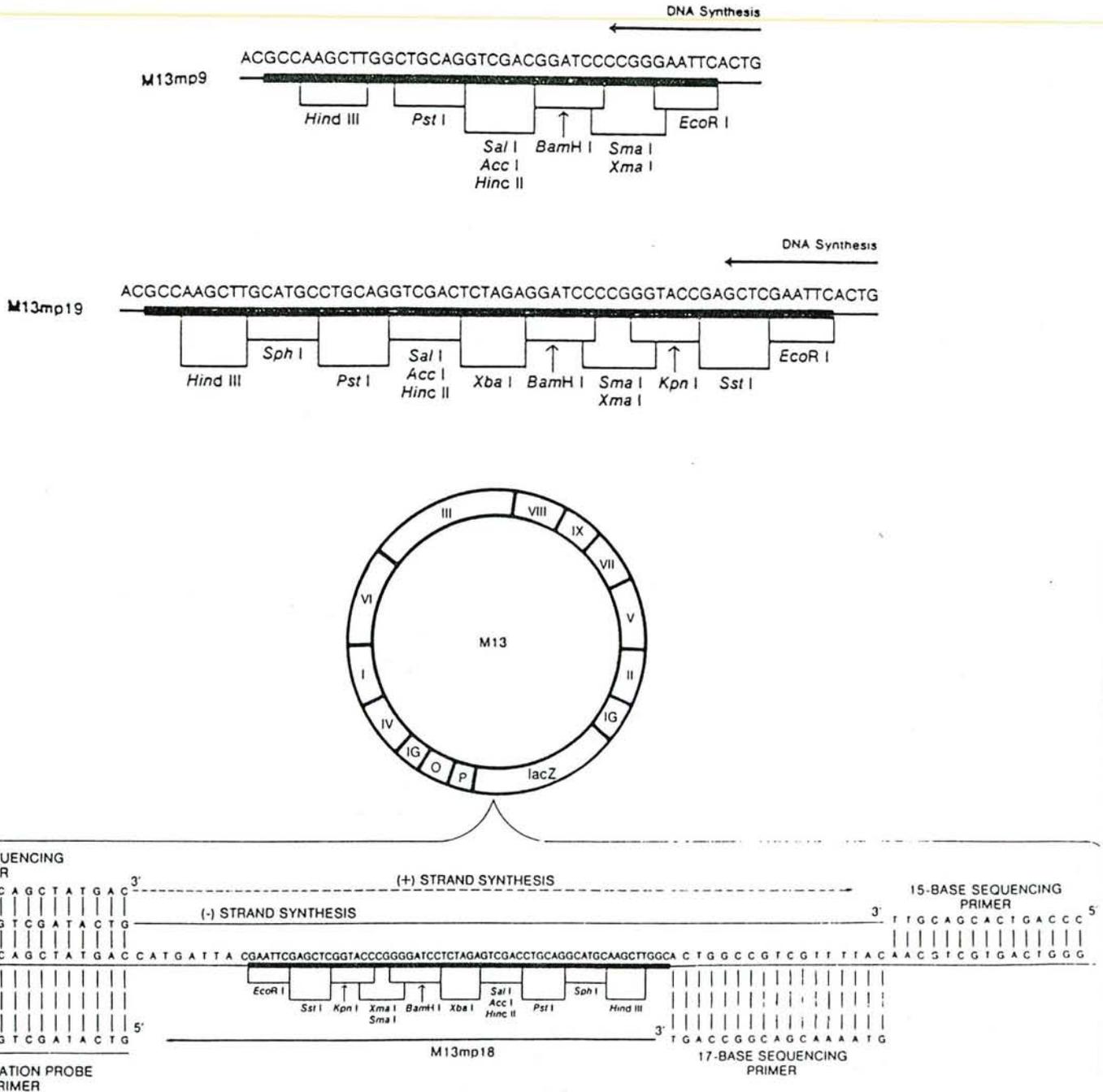
Nous avons utilisé les phages M13mp9, M13mp18 et M13mp19 obtenus après différentes modifications du phage M13 sauvage (Messing et al., 1977). Ces phages renferment en effet la même cassette que les plasmides pUC permettant une insertion dans le fragment du gène de la β -galactosidase. La séquence "polylinker" est plus courte dans M13mp9 et dans la même orientation dans M13mp9 que dans M13mp19 (Fig.17). Les phages M13mp18 et M13mp19 se distinguent de M13 mp9 par l'absence des deux mutations ambre dans les gènes C1 et C3, ce qui justifie leur utilisation lors de la mutagénèse dirigée.



All Bluescript Vectors are 2959 nucleotides

Figure 16 : Cartographie des plasmides Bluescript

Figure 17 : Le phage M13 et ses dérivés



III - TRANSFORMATION DES SOUCHES D'*E.coli*

1. Préparation de cellules compétentes

1-1. Souches JM101, BJ5183, HB101 et HB2154

20 ml de milieu LB sontensemencés au 1/100^e à partir d'une préculture saturée. La culture est placée à 37°C sous agitation.

Lorsque la DO à 700 nm atteint 0,5 (après environ 2 h 30 de culture), les cellules sont centrifugées 15 mn à 4.000 tr/mn (centrifugeuse Beckman J6). Le culot est repris dans 10 ml de CaCl₂ 50 mM stérile et incubé 15 mn dans la glace fondante.

Les cellules sont à nouveau centrifugées et reprises dans 2 ml de cette même solution de CaCl₂. Elles peuvent être utilisées immédiatement, mais leur compétence est optimale après 4 h d'incubation dans la glace fondante. Les cellules conservées à 4°C restent suffisamment compétentes pendant trois jours environ.

1-2. Souche DF221

40 ml de milieu M63 succinate-glycérol sontensemencés au 1/50^e avec une préculture d'une vingtaine d'heures. La culture est placée à 37°C sous agitation pendant environ 3 h 30, temps nécessaire pour que la DO₇₀₀ atteigne 0,32.

Les cellules sont centrifugées 15 mn à 4.000 tr/mn, puis reprises dans 20 ml de la solution suivante :

CH ₃ COONa pH 5,5.....	10	mM
CaCl ₂	10	mM
MnCl ₂	100	mM

Après une incubation de 30 mn à 0°C, les cellules sont centrifugées à nouveau. Le culot est repris dans 2 ml de la solution ci-dessus. Les cellules gardent leur compétence durant trois jours, conservées à 4°C.

2. Transformation de cellules compétentes

2-1. Par les plasmides

2-1-1. Souches BJ5183, HB101 et JM101

Dans un microtube, 20 ng de plasmides sont ajoutés à 100 μ l de la solution de cellules compétentes. Le mélange est placé à 0°C pendant 30 mn. On réalise alors un choc thermique en portant le mélange à 42°C pendant 2 mn. L'ensemble est alors étalé sur milieu LB ampicilline sur lequel se développeront sélectivement les colonies de cellules ayant été transformées par les plasmides. L'incubation dure une nuit à 37°C.

La sélection des colonies transformées par les plasmides est basée sur la résistance aux antibiotiques (ampicilline, tétracycline). Les colonies transformées par les plasmides pUC et Bluescript recombinants peuvent également être sélectionnées en fonction de la présence d'activité β -galactosidase. Cette sélection peut être réalisée dans la souche JM101 en ajoutant X-Gal (5-bromo, 4-chloro, 3-indolyl β -D-galactopyranoside) et IPTG (isopropyl β -D-thiogalactoside) avant d'étaler les cellules transformées (voir paragraphe 2-2.).

Le fragment porté par le plasmide pUC natif complète le gène chromosomique muté (Δ lac) et permet une production de β -galactosidase, responsable de la coloration bleutée des colonies en présence d'IPTG, inducteur de l'expression du gène, et de X-gal, analogue de substrat dont l'hydrolyse conduit à l'apparition de colonies bleues. Le segment "polylinker" est situé dans le fragment codant pour la partie N-terminale de la β -galactosidase, permettant l' α -complémentation de la souche. L'insertion d'un fragment de DNA empêchera l' α -complémentation et entraînera une absence de la coloration.

2-1-2. Souche DF221

Dans un microtube, 100 μ l de cellules compétentes sont mélangés à 100 ng de plasmides. Après incubation à 0°C pendant 45 mn, le mélange est porté à 42°C pendant 2 mn. Le microtube est alors placé à 37°C pendant une heure. Les cellules sont alors centrifugées et reprises dans 100 μ l de milieu M63 succinate-glycérol. Ceci a pour but d'éliminer le chlorure de manganèse qui précipite sur les boîtes. Les cellules sont étalées sur milieu M63 succinate-glycérol-ampicilline ou M63 glucose-ampicilline si l'objectif est de compléter la souche DF221. Une incubation de 24 h à 37°C est nécessaire avant de prélever les colonies.

2-2. Par les phages

Dans un microtube, 100 µl de cellules compétentes sont mélangés à 2 ng de DNA double brin ou 20 ng de DNA simple brin. Le mélange est incubé 30 mn à 0°C et porté à 42°C pendant 2 mn. On ajoute alors 20 µl d'une solution d'IPTG 100 mM et 30 µl d'une solution de X-Gal à 2 % (p/v). L'ensemble est mélangé à 3 ml de gélose molle dans lesquels on ajoute 100 µl d'une culture saturée dont les cellules constitueront un tapis cellulaire permettant un repérage plus facile des plages de croissance ralentie.

Cette culture saturée est composée de cellules de la même souche que les cellules compétentes, excepté dans le cas de la mutagenèse dirigée où les cellules compétentes sont préparées à partir de la souche HB2154 et les cellules du tapis à partir de la souche HB2151.

Le mélange est immédiatement coulé sur milieu LB. Après que ce mélange ait pris en masse, les boîtes sont placées pour une incubation à 37°C pendant la nuit.

IV - PREPARATION DE DNA

1. Préparation du DNA génomique de *C.pasteurianum*

Les cellules de *C.pasteurianum* sont récupérées à partir d'une culture en début de phase exponentielle de croissance. Elles sont centrifugées à 4.000 tr/mn pendant 20 mn. Elles sont lavées dans le tampon A [saccharose 25 % (p/v); Tris-HCl 50 mM pH 7,5 ; EDTA 5 mM] à deux reprises. Le culot est ensuite congelé à -20°C.

Un gramme de cellules est repris dans 10 ml de tampon A. Les cellules sont lysées par 60 mg de lysozyme (en solution à 10 mg/ml dans le tampon Tris-HCl 25 mM pH 8,0) pendant 90 mn à 37°C sous agitation douce. Puis sont ajoutés 4 ml d'EDTA 0,5 M pendant 5 mn à température ambiante. La lyse est achevée en ajoutant 30ml de tampon B [Triton X-100 0,2 % (v/v) ; Tris-HCl 50 mM ; EDTA 50 mM] et en incubant l'ensemble 1 h dans la glace fondante. Les débris cellulaires sont éliminés par centrifugation de 30 mn à 4°C à 15.000 tr/mn (centrifugeuse Beckman J21). Le lysozyme a pour rôle de perméabiliser la paroi bactérienne. L'EDTA facilite cette action en chélatant les ions Ca²⁺ de la paroi. Le Triton X-100 dissout les lipides membranaires et dénature les protéines.

Le surnageant est récupéré. Du chlorure de césium y est dissous à raison de 0,9 g/ml ainsi que du bromure d'éthidium à 0,02 % (p/v) final. L'ensemble est homogénéisé et réparti dans des tubes à scellement rapide. Après une centrifugation en rotor vertical à 60.000 tr/mn (centrifugeuse Beckman L8-M) pendant 6 h à 20°C, un gradient de densité de CsCl s'établit, permettant de séparer les acides nucléiques (DNA génomique, DNA plasmidique ou phagique éventuellement et RNA). Ils sont rendus visibles sous U.V. par fluorescence du BrEt intercalé.

La bande de DNA génomique est prélevée à la seringue. Le BrEt est extrait à l'isopropanol (vol/vol). L'extraction est achevée par un traitement au butanol saturé dans le tampon TE [Tris-HCl 10 mM ; EDTA 1 mM ; pH 8,0] après avoir dilué la solution initiale trois fois dans l'eau. Le butanol est éliminé par centrifugation, puis évaporation dans un dessiccateur pendant 2 h environ. La solution obtenue est diluée 10 fois au total et le DNA est sédimenté par centrifugation à 60.000 tr/mn pendant 15 h à 4°C. Le culot est séché et repris dans 200 µl de solution [Tris 10 mM; NaCl 50 mM ; pH 8,0].

Les protéines fixées au DNA génomique sont extraites par deux traitements successifs au phénol et au chloroforme. Le DNA est alors précipité à l'éthanol et repris dans 200 µl d'une solution [Tris-HCl 10 mM pH 7,5 ; NaCl 50mM ; MgCl₂ 10 mM]. Le dosage de la quantité de DNA ainsi obtenue est réalisé par spectrophotométrie U.V. à 260 nm. Il est conservé à 4°C.

2. Préparation du DNA génomique de *B.stearothermophilus*

Par rapport au protocole de préparation ci-dessus, nous avons utilisé des conditions de lyse plus douces, en raison de la moindre résistance offerte par la paroi des cellules de *B.stearothermophilus*.

La lyse est réalisée à 0°C pendant 10 mn avec 50 mg de lysozyme par gramme de cellules. Puis sont ajoutés 5 ml d'EDTA 0,5 M pour 5 mn à 0°C. La lyse est terminée par une incubation de 1 h à 0°C avec 30 ml de tampon B. Les étapes ultérieures sont identiques à celles utilisées pour la préparation du DNA génomique de *C.pasteurianum*.

3. Préparation de DNA plasmidique

3-1. Technique du lysat clair

Une culture de cellules transformées est effectuée dans un litre de milieu LB ampicilline et amenée en phase stationnaire. Les cellules sont centrifugées à 4.000 tr/mn pendant 20 mn. Le culot est lavé dans le tampon A, centrifugé à nouveau pendant 30 mn et congelé à -20°C.

Pour la lyse, le culot est repris dans 20 ml de tampon A. Les cellules sont lysées avec 50 mg de lysozyme pendant 8 mn à 0°C. Après l'ajout de 5 ml d'EDTA 0,5 M, une incubation est pratiquée 5 mn à 0°C. La lyse est complétée par l'ajout de 30 ml de tampon B et une incubation du mélange, tout d'abord 1 h à 0°C, puis 15 mn à 45°C.

Après centrifugation de 30 mn à 20.000 tr/mn, le surnageant est recueilli et additionné de 1/3 de volume de PEG 6.000 à 40 % (v/v) et de 1/4 de volume de NaCl 5 M. Cette étape permet de concentrer et précipiter le DNA plasmidique. L'incubation est menée 1 h à 0°C.

Le précipité est obtenu après centrifugation de 15 mn à 20.000 tr/mn. Il est repris dans 18 ml de TE contenant 18 g de CsCl et 0,02 % (p/v) de BrEt. Après homogénéisation, le mélange est réparti en quatre tubes à scellement rapide et centrifugé 6 h à 20°C à 60.000 tr/mn. La bande inférieure composée de DNA plasmidique est prélevée à la seringue. Le BrEt est extrait à l'isopropanol, puis au butanol. Le DNA plasmidique est alors traité au phénol et au chloroforme, précipité à l'éthanol, repris dans du tampon TE stérile et conservé à -20°C.

3-2. Technique de lyse alcaline (Birnboim et Doly, 1979)

Cette technique est destinée à la préparation rapide de petites quantités de plasmides. Elle est réalisée en microtubes à partir de cultures de 1 à 2 ml. Les bactéries sont centrifugées puis lysées dans 100 µl de solution I [glucose 50 mM ; Tris-HCl (pH 8,0) 25 mM ; EDTA 10 mM] dans laquelle est ajouté extemporanément le lysozyme à raison de 4 mg/ml. L'incubation est réalisée 5 mn à température ambiante. La solution II [NaOH 0,2 N ; SDS 1 % (p/v)] est ajoutée alors à raison de 200 µl, le tout est incubé 5 mn à 0°C. La lyse est achevée en neutralisant le mélange par 150 µl d'une solution d'acétate de potassium 5 M à pH 4,8, puis en incubant 5 mn à 0°C. Le surnageant est prélevé après centrifugation. Après deux extractions au phénol puis au chloroforme, le DNA est précipité lors d'une incubation de 2 mn après avoir ajouté 1ml d'éthanol à température ambiante. Le précipité est séché après centrifugation et lavage à l'éthanol 70 % (v/v). Il est repris

dans 50 µl de tampon TE contenant 20 µg/ml de RNase pancréatique sans DNase. Après 1 h d'incubation à 37°C, le DNA est conservé à -20°C ; il peut être utilisé pour les digestions enzymatiques.

4. Préparation du DNA du phage M13

4-1. Rappels sur la physiologie du phage M13

M13 est un phage filamenteux contenant un DNA monocaténaire circulaire. En infectant les cellules-hôtes présentant le phénotype F', le phage abandonne son enveloppe protéique. Il ne lyse pas les bactéries qu'il infecte, mais ralentit leur croissance deux à trois fois. Le DNA viral simple brin est alors converti en une molécule bicaténaire, la forme répliquative, qui est répliquée à plus d'une centaine de copies. A partir de celles-ci sont synthétisées de nouvelles molécules de DNA viral monocaténaires. Ces molécules (+) infectieuses sont alors encapsidées et excrétées en dehors de la cellule (Fig. 18).

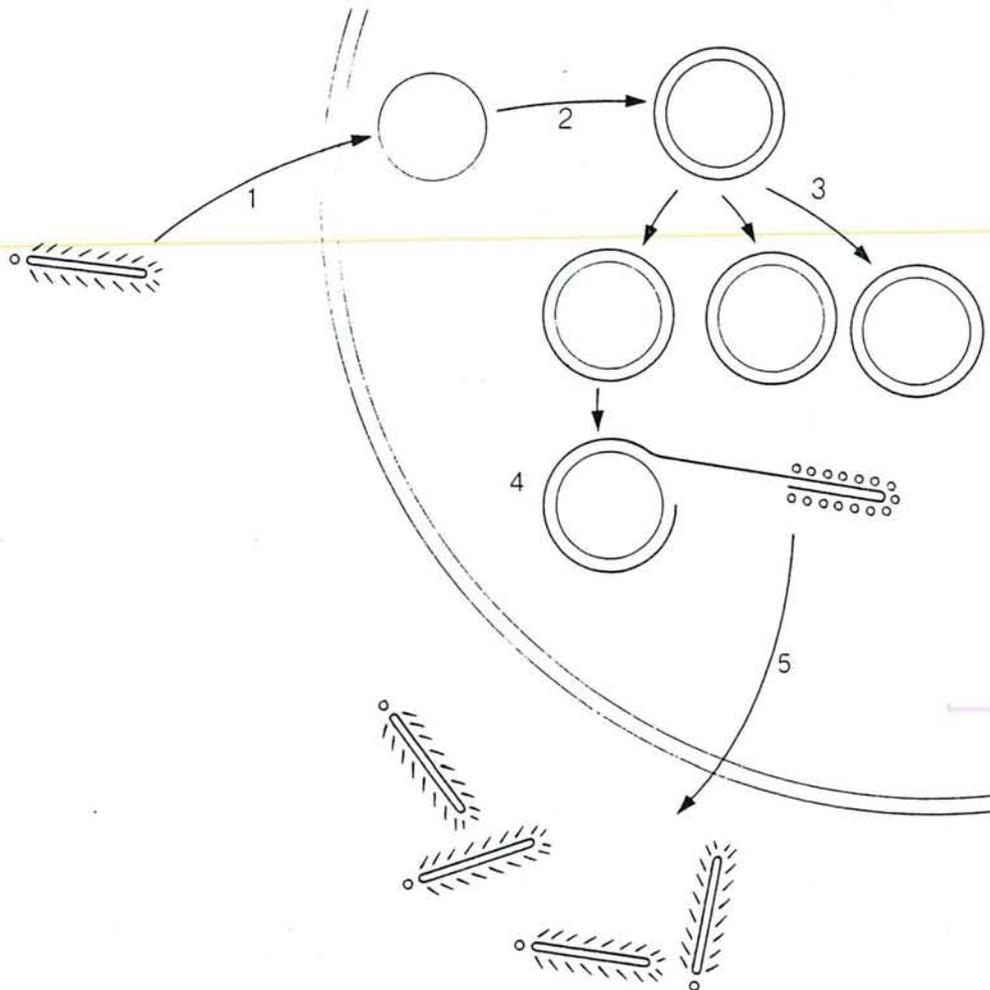
La préparation de DNA du phage M13 sera donc différente selon que l'on désire obtenir la forme infectieuse simple brin (sb) ou la forme répliquative double brin (db).

4-2. Forme simple brin (sb)

Des cellules d'*E.coli* JM101 ou HB2154 infectées par le phage M13 ont été étalées après transformation sur une boîte de Pétri. A l'aide d'un cure-dent stérile, une plage est prélevée et utilisée pour ensemercer 1,5 ml de milieu LB en tube.

Après 6 h de culture à 37°C sous agitation, les cellules sont éliminées par centrifugation à 13.000 tr/mn pendant 15 mn. Le surnageant est récupéré. Les phages sont précipités en présence de 200 µl de [PEG 6.000 20 % (v/v) ; NaCl 2,5 M] pendant 20 mn à température ambiante. Une centrifugation de 20 mn à 13.000 tr/mn est réalisée et le surnageant est soigneusement éliminé. Le précipité est repris dans 50 µl de tampon TE. Après deux extractions au phénol et au chloroforme visant notamment à éliminer les protéines de la capsule, le DNA est précipité à l'éthanol et repris dans 20 µl de tampon TE stérile. Il est conservé à -20°C.

Figure 18 : Le cycle de vie du phage M13



1. Le phage pénètre dans la bactérie-hôte (*E. coli F'*) par son facteur F pilus.
2. Le DNA sb est converti en forme db, la forme réplivative.
3. Par réplivation, la forme réplivative se multiplie (plus de 100 copies).
4. Synthèse de phage simple brin par le procédé du cercle tournant.
5. Le DNA sb est encapsidé et excrété de la cellule.

4-3. **Forme double brin (db)**

A partir d'une plage de JM101 infectée par le phage M13, 2,5 ml de milieu LB sontensemencés pour une préculture de 6 h à 37°C sous agitation. Une heure après, une culture de JM101 estensemencée, à partir d'une préculture de la veille, dans 2,5 ml de milieu LB. Cette culture est arrêtée après 2 h 30 environ, à une DO₇₀₀ de 0,5, et est utilisée pour ensemen- cer 250 ml de milieu LB qui sont placés à 37°C sous agitation. Après 2 h 30, la préculture de JM101 infectée par le phage M13 y est ajoutée pour 4 h supplémentaires de culture.

Les cellules sont alors centrifugées et lavées dans le tampon A. La forme double brin du phage M13 est alors préparée selon la technique du lysat clair utilisée pour préparer les plasmides.

5. Purification du DNA

5-1. **Extractions au phénol et au chloroforme**

Le rapport de concentrations acides nucléiques/protéines est estimé par mesure spectrophotométrique des densités optiques 260 nm/280 nm. Un DNA parfaitement purifié présente un rapport de 2. Nous avons essayé d'obtenir un rapport toujours égal ou supérieur à 1,7. Lorsque le rapport était inférieur à cette valeur, le DNA était soumis à une ou plusieurs extractions phénoliques et chloroformiques qui permettent d'éliminer les protéines encore fixées sur le DNA. Ce type d'extraction a également été utilisé pour éliminer les enzymes d'un milieu d'hydrolyse.

Cette méthode consiste à ajouter à la solution contenant le DNA un volume égal de phénol saturé ou de chloroforme contenant 1/24^e d'alcool isoamylique. Après agitation et centrifugation de 3 mn à 13.000 tr/mn, la phase aqueuse est prélevée.

5-2. **Précipitation à l'éthanol**

Le DNA est précipité dans trois volumes d'éthanol froid contenant du NaCl à la concentration de 0,3 M. Le mélange est homogénéisé et placé 10 mn à -80°C ou une nuit à -20°C. Le surnageant est éliminé après centrifugation de 20 mn à 13.000 tr/mn à 0°C. Le culot est lavé de ses sels dans l'éthanol à 70 % (v/v), centrifugé à nouveau, séché et repris dans du tampon TE.

V - DIGESTION, DEPHOSPHORYLATION, SEPARATION ET LIGATION DE DNA

1. Digestion par les enzymes de restriction

Une unité d'enzyme (U), représente la quantité nécessaire à la digestion d'un microgramme de DNA du bactériophage λ pendant 1 h à 37°C. Dans la plupart des cas, nous ne nous sommes pas contentés d'utiliser les conditions définies par le fournisseur. Nous avons en effet réalisé des cinétiques d'hydrolyse, afin de définir les meilleures conditions d'utilisation. Dans tous les cas, nous avons essayé de toujours nous placer à une concentration en substrat de 1 $\mu\text{g}/10 \mu\text{l}$.

Enzyme	Site de restriction 5' vers 3'	Force ionique
AluI	AG CT	Moyenne
BamHI	G GATCC	Forte
DraI	TTT AAA	Faible
HaeIII	GG CC	Moyenne
HindII	GTPy PuAC	Moyenne
HindIII	A AGCTT	Moyenne
EcoRI	G AATTC	Forte
PstI	CTCGA G	Moyenne
Sau3A	GATC	Moyenne

Les tampons à faible, moyenne et forte force ionique sont les suivants :

Force ionique	Tris-HCl (pH 7,5)	MgCl ₂	NaCl	DTT
Faible	10 mM	10 mM	-	1 mM
Moyenne	10 mM	10 mM	50 mM	1 mM
Forte	50 mM	10 mM	100 mM	1 mM

Certaines enzymes nécessitent des tampons plus spécifiques.

Enzyme	Site	Tris-HCl	MgCl ₂	NaCl	DTT	β-ME
EcoRV	GTT AAC	6 mM (pH 7,9)	6 mM	150 mM	-	6 mM
NdeI	CA TATG	10 mM (pH 7,5)	6 mM	150 mM	1 mM	-
Sall	G TCGAC	6 mM (pH 7,9)	6 mM	150 mM	-	6 mM
ScaI	AGT ACT	6 mM (pH 7,4)	6 mM	100 mM	-	6 mM

2. Déphosphorylation du DNA

Cette étape, nécessaire dans le protocole de préparation d'un vecteur, consiste à éliminer le groupement phosphate situé à l'extrémité 5' d'une molécule de DNA hydrolysée, ce qui l'empêche de se religuer sur elle-même.

Cette opération est réalisée à raison de 0,4 U de phosphatase alcaline d'intestin de veau par picomole d'extrémité 5' libre, pour 1 h au moins à 37°C.

3. Séparation électrophorétique des fragments générés

3-1. En gel d'agarose

L'électrophorèse horizontale en gel d'agarose constitue une méthode simple et rapide pour séparer, identifier ou purifier des fragments de DNA db ou sb. Les gels que nous avons généralement utilisés contiennent 0,8 % (p/v) d'agarose dans du tampon TEA [Tris-HCl 20 mM ; EDTA 1 mM ; CH₃COONa 10 mM] et 100 µg de BrEt.

Dans ce type de gel, la quantité minimale visualisable est de 200 ng environ. L'échantillon est additionné de 1/2 volume de solution de dépôt [EDTA 125 mM ; SDS 2 % (p/v) ; glycérol 25 % (v/v) ; bleu de bromophénol 0,12 % (p/v)] et déposé dans un puits. La migration est réalisée à 80 V pendant 2 à 3 h. Le DNA est visualisé sous U.V. par fluorescence du BrEt intercalé.

En vue de leur purification, des fragments ont été séparés par cette même technique d'électrophorèse, mais en gel d'agarose-NA qui présente une qualité supérieure.

3-2. En gel de polyacrylamide

L'électrophorèse en gel de polyacrylamide a été utilisée pour fractionner des fragments de DNA de taille inférieure à 800 pb. La concentration du gel étant adaptée à la taille des fragments à purifier, nous avons généralement utilisé des gels à 6 % (p/v) qui permettent de séparer correctement des fragments de 60 à 500 pb.

Les gels de polyacrylamide sont obtenus par polymérisation de deux monomères, l'acrylamide et le N, N'-méthylène bisacrylamide, induite par deux catalyseurs, le N, N, N', N'-tétraméthyl éthylène diamide (TEMED) et le persulfate d'ammonium.

Les gels utilisés ont une hauteur de migration de 400 mm, une largeur de 200 ou 300 mm et une épaisseur de 0,5 ou 1 mm.

Pour constituer un gel de fractionnement à 6 % (p/v) de polyacrylamide et de dimensions 200 mm x 400 mm x 1 mm, la solution suivante est préparée :

- 15 ml solution stock désionisée acryl/bisacryl 40/1,36 % (p/v) ;
- 5 ml tampon TEB 20x [Tris 2 M ; EDTA 50 mM ; acide borique 1,76 M ;
pH 8,3] ;
- 100 µl TEMED 6,67 M ;
- H₂O ultrapure qsp 100 ml.

Le gel est polymérisé en ajoutant à ce mélange 700 µl de la solution de persulfate d'ammonium à 100 mg/ml. Les puits sont ensuite lavés soigneusement à l'aide de tampon TEB.

Les échantillons sont déposés après que leur ait été ajouté 1/2 volume de solution de dépôt. La migration est effectuée pendant une nuit sous un voltage de 250 V, c'est-à-dire de 60 V/cm, dans du tampon TEB.

Le gel est ensuite placé dans un bac rempli de ce tampon contenant du BrEt à la concentration de 1 mg/l. Après 30 mn, le DNA peut être visualisé par fluorescence sous U.V. du BrEt intercalé.

4. Electroélution des fragments séparés

Les bandes de gel contenant les fragments de DNA intéressants sont prélevées et introduites séparément dans des sacs à dialyse contenant du tampon TEB et fermés par des pinces. Les sacs sont placés perpendiculairement au champ électrique sur un appareil horizontal. L'électroélution, menée à 80 V pendant 3 h, concentre le DNA sur la paroi du sac. L'inversion de la polarité du champ électrique pendant quelques instants permet de "détacher" le DNA de la paroi. Il est alors récupéré en solution dans un microtube, précipité à l'éthanol et repris dans du tampon TE stérile. Lorsque la quantité de DNA à purifier est très faible, 1 µg de tRNA est ajouté pour servir d'entraîneur lors de la précipitation.

5. Ligation

Cette étape permet de religuer des molécules de DNA sur elles-mêmes ou d'insérer des fragments de DNA dans un vecteur présentant des extrémités compatibles.

La ligation est réalisée par la DNA ligase du phage T4 qui catalyse la formation de liaisons phosphodiester entre les groupements phosphate en 5' et hydroxyle en 3' du DNA. Cette réaction est ATP-dépendante.

Le mélange réactionnel possède la composition suivante :

- 100 ng vecteur coupé et phosphatasé ;
- 1 µl ATP 10 mM ;
- 1 µl DTT 100 mM ;
- 1 µl tampon de ligation 10x [Tris-HCl 0,2 M ; MgCl₂ 0,1 M] ;
- DNA à insérer (quantité variable selon la longueur du fragment) ;
- H₂O qsp 10 µl.

L'incubation s'effectue à 15°C pendant une nuit pour les ligations de fragments à bouts francs et à température ambiante pendant 4 h pour celles de fragments à extrémités cohésives.

Les conditions énoncées ci-dessus peuvent être modifiées en fonction de l'objectif recherché, notamment les ligations intramoléculaires favorisées par une grande dilution du milieu réactionnel.

Le mélange de ligation est utilisé directement pour transformer les cellules compétentes d'*E.coli*.

VI - PREPARATION DE DEOXY-OLIGONUCLEOTIDES DE SYNTHÈSE

Utilisés pour le séquençage ou la mutagenèse dirigée, les oligonucléotides doivent être choisis pour qu'ils s'hybrident, facilement et spécifiquement, uniquement sur la région concernée.

Les oligonucléotides ont été synthétisés automatiquement sur l'appareil "Gene Assembler" de Pharmacia.

1. Synthèse chimique

1-1. Principes généraux (Gait, 1984)

La synthèse chimique d'oligonucléotides consiste en la condensation du groupement hydroxyle en 5' du désoxyribose du dernier nucléotide avec le groupement phosphate en 3' du nucléotide suivant. Des groupements protecteurs sont greffés sur certaines fonctions réactives afin d'éviter les réactions secondaires qui affecteraient le rendement de la synthèse. Ils doivent satisfaire à trois conditions indispensables :

- bloquer les sites réactifs ;
- rester stables dans les conditions imposées lors du couplage ;
- s'éliminer facilement en fin de réaction sans occasionner de dégradation de la molécule synthétisée.

1-2. Eléments de base

La synthèse réalisée sur l'appareil "Gene Assembler" est menée selon la méthode aux phosphites utilisant comme molécules de base les cyanoéthyl-phosphoramidites qui possèdent un phosphore trivalent extrêmement réactif. Dans cet état, l'atome de phosphore, déficient en électrons, réagit facilement avec tout groupement donneur, ce qui explique la nécessité de conditions de stockage et de synthèse totalement anhydres, l'eau elle-même pouvant jouer le rôle de donneur d'électrons.

Les phosphoramidites sont introduits dans la réaction avec un certain nombre de groupements protégeant les fonctions réactives de ces éléments de base.

1-2-1. Le groupement hydroxyle en 5' du désoxyribose

Le radical 4, 4'-diméthoxytrityl (DMTr) est utilisé pour protéger le groupement hydroxyle en 5' de chaque nucléotide ajouté qui pourrait entrer en compétition lors de la réaction de couplage avec l'hydroxyle en 5' du dernier résidu de la chaîne en élongation. Ce radical est facilement éliminé par détritylation en milieu acide, préparant la chaîne au couplage du nucléotide suivant.

1-2-2. Les bases

Les groupements amines des bases sont protégés spécifiquement : l'adénine par le chlorure d'acide benzoïque, la cytosine par l'anhydride benzoïque et la guanine par le chlorure d'acide isobutyrique. La thymine ne possède pas de groupements réactifs à protéger.

Ces groupements protecteurs sont aisément éliminés en fin de synthèse par traitement à l'ammoniaque concentrée.

1-2-3. Le groupement phosphate en 3' du désoxyribose

Ce groupement est protégé par le radical β -cyanoéthyle ($-\text{CH}_2-\text{CH}_2-\text{CN}$). Ce substituant aussi est facilement éliminé lors de l'étape de déprotection en fin de synthèse.

1-3. La synthèse

1-3-1. Le support solide

La synthèse d'oligonucléotides en phase solide nécessite un support insoluble, inerte et stable dans les solvants utilisés, résistant aux pressions appliquées durant la synthèse et présentant une porosité suffisante pour permettre les réactions intermoléculaires.

Le procédé "Gene Assembler" utilise un support constitué de billes de résine synthétique de 10 μm de diamètre. Cette résine est fournie par Pharmacia avec le premier nucléoside fixé par un bras espaceur. Ce nucléoside correspond au nucléotide qui sera situé à l'extrémité 3' de la molécule. L'oligonucléotide est donc synthétisé dans le sens 3' vers 5'.

1-3-2. Les réactions

La synthèse est basée sur la très grande réactivité des phosphites.

Sur le groupement phosphite du phosphoramidite se trouve fixée une amine secondaire, la diisopropylamine. Ce groupement fonctionnel n'est pas réactif en tant que tel avec les groupements hydroxyle, mais il est facilement activé par un acide faible comme le tétrazole, composé intermédiaire qui réagit avec le groupement hydroxyle en 5' d'un déoxyribonucléotide préalablement détritylé. L'efficacité de cette réaction de couplage dépasse 98 % en quelques minutes. Le solvant anhydre utilisé pour cette réaction est l'acétonitrile.

La synthèse d'un oligonucléotide consiste donc en une succession de cycles (Fig. 19) dont les étapes sont chronologiquement :

- détritulation, par l'acide dichloroacétique, de l'hydroxyle en 5' du dernier nucléotide fixé à la chaîne ;
- activation du phosphite triester en 3' du nucléotide à incorporer ;
- couplage ;
- oxydation du phosphite en phosphate, beaucoup plus stable ;
- blocage (ou "capping") des fonctions hydroxyle en 5' des nucléotides qui n'ont pas été couplés.

Cette dernière étape est réalisée par une solution d'anhydride acétique à 20 % (v/v) et de diméthylaminopyridine à 6 % (v/v) dans l'acétonitrile. Elle consiste en l'acylation de l'oxygène des fonctions hydroxyle en 5' en acétates.

En fin de synthèse, la cassette contenant l'oligonucléotide fixé sur le support est placée dans un microtube. L'acétonitrile adsorbé à la résine est éliminé par une courte centrifugation à 2.000 tr/mn et évaporation en "speed vac". La cassette peut alors être conservée à -20°C.

2. Déprotection et purification de l'oligonucléotide

Après sa synthèse, l'oligonucléotide est fixé à la résine, tritylé en 5', les bases et les phosphates sont protégés. Il est alors nécessaire de déprotéger puis de purifier la molécule synthétisée.

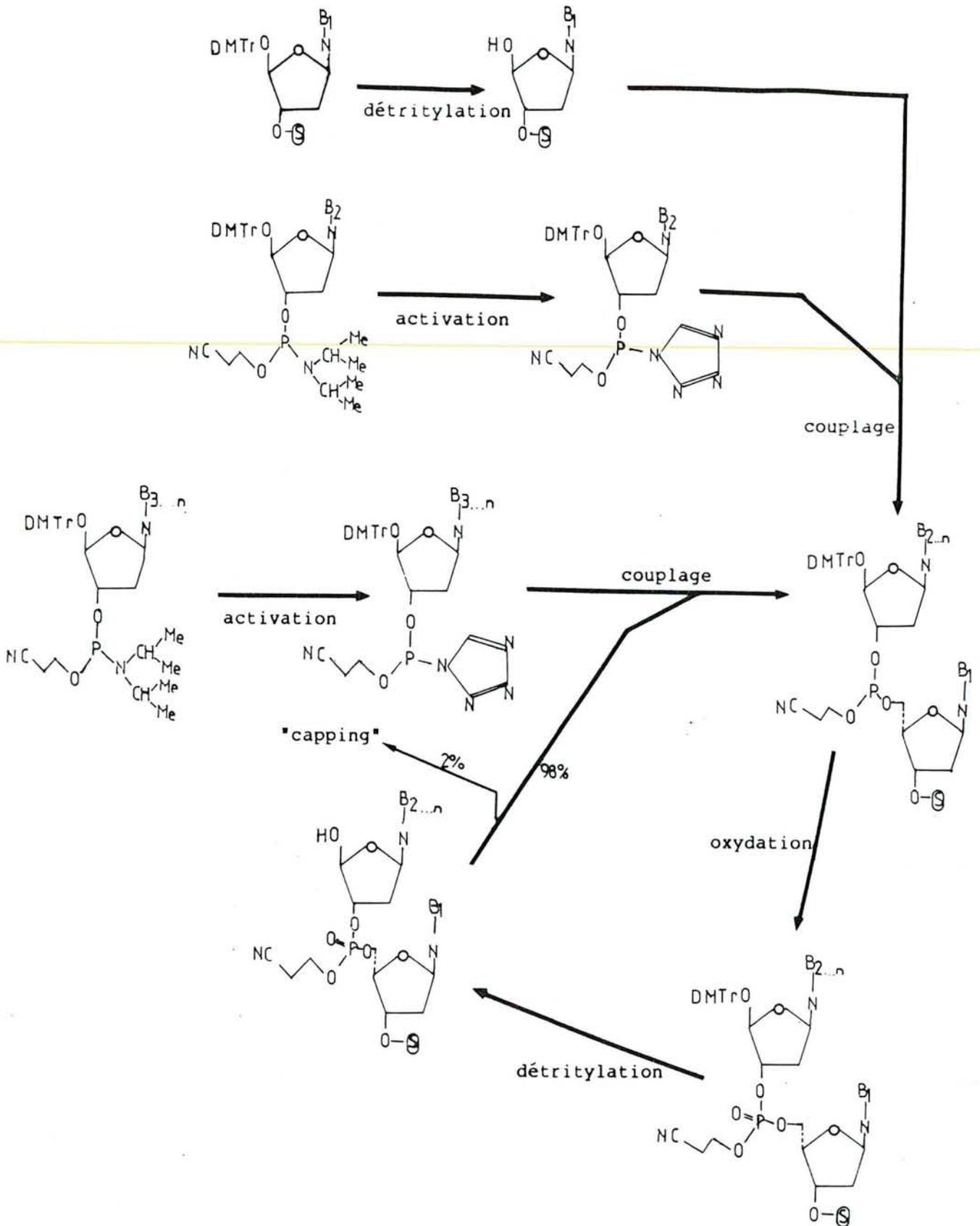


Figure 19 : Synthèse chimique des oligonucléotides.
Méthode aux phosphites.

2-1. Solubilisation et déprotection

Cette étape consiste en l'hydrolyse des liaisons amides protégeant les bases et les phosphates. Elle permet simultanément de solubiliser l'oligonucléotide en rompant la liaison ester-succinate établie entre l'hydroxyle en 3' du premier nucléotide et le bras succinate fixé sur le support. Seul, le groupement DMTr reste fixé puisqu'il n'est pas labile en conditions alcalines.

La cassette est transférée dans un tube à vis de 2 ml dans lequel est ajouté 1 ml de NH_4OH concentrée à 30 % (v/v). Le tube est placé à 50°C pendant 16 h.

En fin d'incubation, l'oligonucléotide déprotégé se trouve dans la solution ammoniacale. L'ammoniaque est évaporée en "speed vac" en trois étapes alternées avec des lavages à l'eau ultrapure. L'oligonucléotide est alors précipité à l'éthanol et repris dans une solution d'acétate de triéthyl-ammonium (TEAA) 0,1 M à pH 7,0.

2-2. Purification par H.P.L.C.

Cette étape est utilisée pour purifier les molécules désirées des chaînes ne présentant pas la taille requise ou détritillées. La purification est menée selon le principe de la chromatographie en phase inverse. La précolonne et la colonne contiennent toutes deux une résine de silice sur laquelle ont été greffées des chaînes en C18, conférant un caractère hydrophobe au support.

La présence du groupement DMTr en 5' augmente l'hydrophobie de la molécule et la retient sur la colonne. Cette hydrophobie croît également en fonction du nombre de nucléotides, mais la composition en bases n'influe pratiquement pas sur la rétention.

La chromatographie est menée dans du tampon TEAA 0,1 M pH 7,0. L'élution est réalisée par un gradient linéaire d'acétonitrile de 5 à 40 % (v/v) dans le tampon TEAA 0,1 M pH 7,0.

Le fractionnement est suivi par un détecteur U.V. à 254 nm. Les fractions renfermant l'oligonucléotide sont collectées, concentrées et lavées dans une centrifugeuse "Speed-Vac". L'oligonucléotide est précipité à l'éthanol et récupéré sous forme de culot séché par évaporation. A ce stade, il est purifié et encore tritylé.

2-3. Détritylation de l'extrémité hydroxyle 5'

L'oligonucléotide est repris dans 100 µl d'acide acétique à 80 % (v/v). Après 20 mn d'incubation à température ambiante, la détritylation est effective, ce qui est visualisé par une coloration rosée ; 200 µl d'eau sont ajoutés à cette solution. L'acide acétique contenant les groupements DMTr est extrait par 800 µl d'éther éthylique en cinq étapes successives.

L'oligonucléotide est finalement précipité à l'éthanol avant d'être repris dans de l'eau ultrapure et dosé à 260 nm. Une aliquote est prélevée pour son utilisation immédiate, tandis que la quantité restante est conservée lyophilisée à -20°C.

VII - SEQUENCAGE DE DNA

1. Technique sur le phage M13 (Sanger et al., 1977)

1-1. Principes

Le fragment de DNA à séquencer est introduit dans l'un des sites de restriction de la séquence "polylinker" du phage M13. Un oligonucléotide de 17-mer qui s'hybride spécifiquement sur le phage simple brin en 3' du fragment à séquencer est utilisé comme amorce au fragment Klenow de la DNA polymérase I. Il définit une origine unique aux fragments synthétisés par l'enzyme. L'élongation réalisée par l'enzyme de Klenow nécessite un mélange des quatre déoxynucléotides. Afin de déterminer la séquence nucléotidique du fragment inséré, quatre mélanges seront constitués, renfermant les quatre déoxynucléotides et l'un des quatre didéoxynucléotides. Les didéoxynucléotides sont dépourvus du groupement hydroxyle en 3' indispensable à l'établissement de la liaison phosphodiester avec le nucléotide suivant (Fig. 20). En s'incorporant, un didéoxynucléotide arrêtera spécifiquement l'élongation de la chaîne (Fig. 21). Le rapport déoxy/didéoxynucléotide est déterminé expérimentalement en fonction de la longueur et de la composition en bases du fragment de DNA à séquencer. Il est choisi de façon à permettre un arrêt statistique de la synthèse des chaînes. L'utilisation d' $[\alpha\text{-}^{32}\text{P}]\text{-dATP}$ permet de marquer les fragments synthétisés. Ils peuvent ainsi être visualisés par autoradiographie.

Figure 20 : Structure d'un déoxynucléotide et d'un didéoxynucléotide

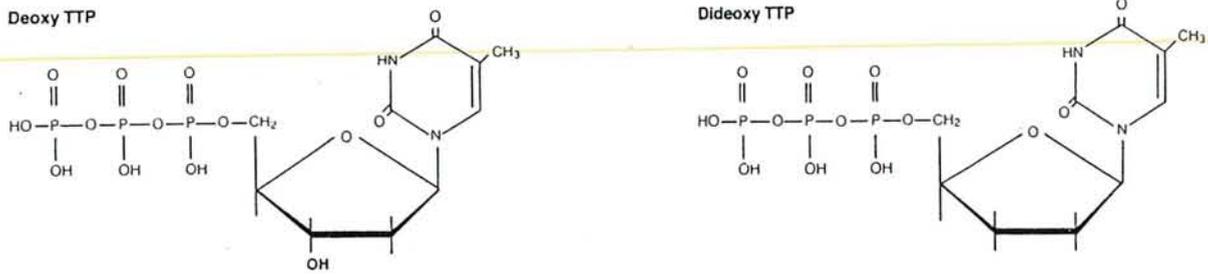
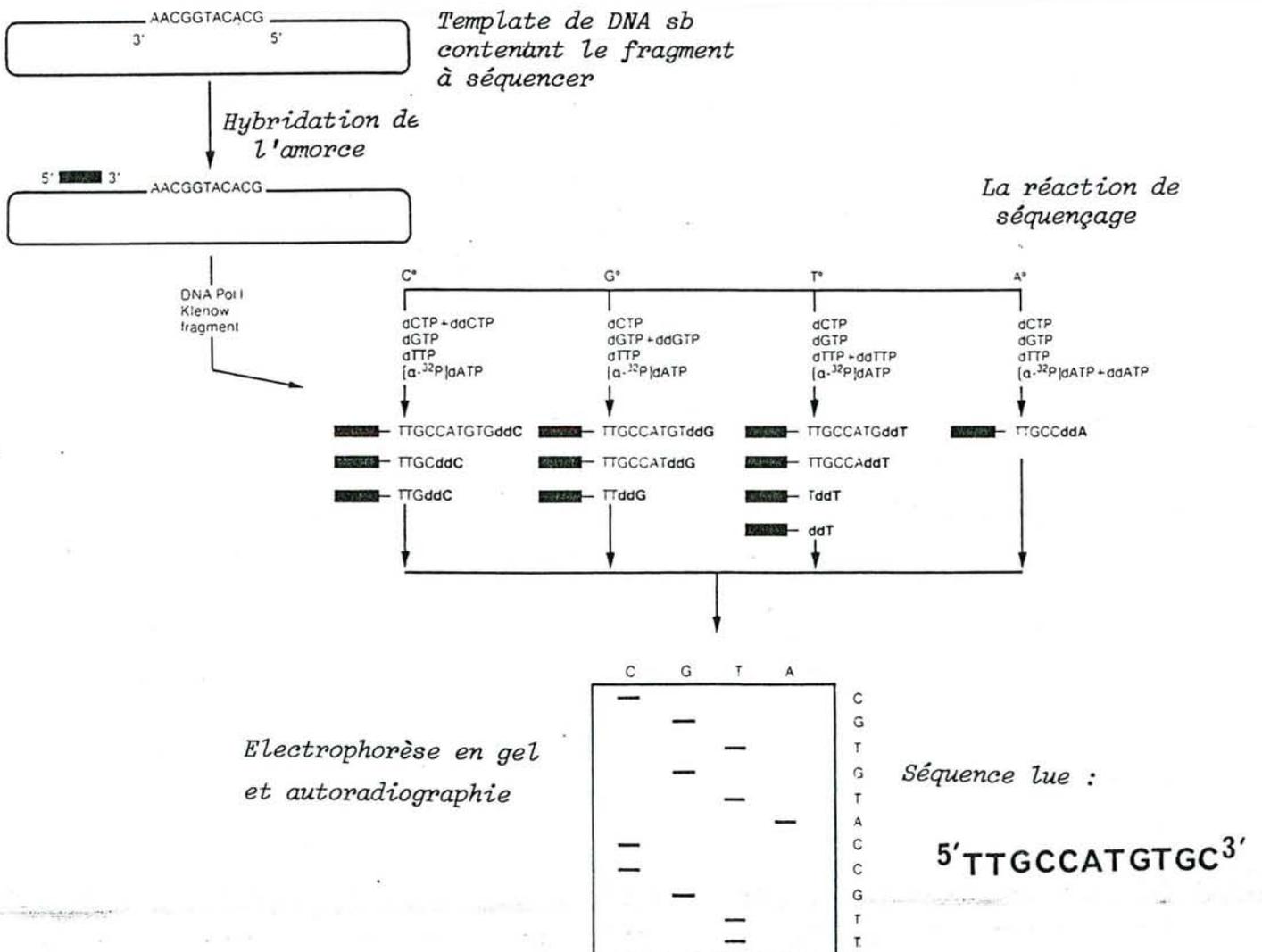


Figure 21 : Principe du séquençage par la méthode aux didéoxynucléotides de Sanger



1-2. Mise en évidence de clones recombinants du phage M13

L'insertion d'un fragment de DNA à extrémités symétriques dans le phage M13 peut s'effectuer dans deux orientations possibles. Avant le séquençage, il est important d'identifier des phages ayant inséré les fragments dans les deux orientations. Pour cela, une hybridation entre les différents phages recombinants obtenus est réalisée en préparant le mélange suivant :

- 2 µl DNA sb du clone a ;
- 2 µl DNA sb du clone b ;
- 6 µl tampon d'hybridation.

Le tampon d'hybridation est préparé extemporanément :

- 1 µl tampon Hin 10x [NaCl 500 mM ; MgCl₂ 100mM; Tris 10 mM] ;
- 1 µl SDS 1 % (p/v) ;
- 1 µl glycérol 50 % (v/v) ;
- 1 µl DTT 100 mM ;
- 0,3 µl bleu de bromophénol ;
- 1,7 µl H₂O.

Pour chaque fragment, six sb différents sont préparés et testés deux à deux pour vérifier la complémentarité des fragments insérés.

L'hybridation est réalisée à 65°C pendant 1 h. Les mélanges sont aussitôt plongés dans la glace fondante et déposés sur un gel d'agarose classique. Après migration et révélation sous U.V., l'hybride formé apparaît sous l'aspect d'une bande supplémentaire, de plus faible mobilité électrophorétique du fait de sa taille.

1-3. Etapes

1-3-1. Solutions nécessaires au séquençage

A partir de solutions concentrées à 10 mM, nous avons préparé des solutions intermédiaires de déoxynucléotides à 0,5 mM dans l'eau ultrapure.

- Les mélanges de déoxynucléotides

	dCTP 0,5 mM	dGTP 0,5 mM	dTTP 0,5 mM	TE
A _o	40 µl	40 µl	40 µl	40 µl
C _o	2 µl	40 µl	40 µl	40 µl
G _o	40 µl	2 µl	40 µl	40 µl
T _o	40 µl	40 µl	2 µl	40 µl

- Les solutions de didéoxynucléotides

- . dd ATP 0,15 mM
- . dd GTP 0,05 mM
- . dd CTP 0,02 mM
- . dd TTP 0,5 mM.

- Les mélanges déoxy/didéoxynucléotides

La composition de chacun de ces mélanges a été optimisée en fonction de la faible proportion de G + C, 26 à 28 %, qui caractérise le génome de *C.pasteurianum*.

- . A_o/ddATP : 15 µl A_o + 5 µl ddATP
- . C_o/ddCTP : 20 µl C_o + 5 µl ddCTP
- . G_o/ddGTP : 15 µl G_o + 5 µl ddGTP
- . T_o/ddTTP : 30 µl T_o + 3 µl ddTTP

- La solution de chasse

Les quatre déoxynucléotides sont présents en quantités non limitantes. La chasse permet de prolonger rapidement la synthèse des chaînes non arrêtées par l'incorporation d'un didéoxynucléotide, afin qu'elles n'interfèrent pas, lors de la lecture, avec celles arrêtées spécifiquement par cet analogue de nucléotide.

La composition de cette solution est la suivante :

- 1 μ l dATP 10 mM ;
- 1 μ l dCTP 10 mM ;
- 1 μ l dGTP 10 mM ;
- 1 μ l dTTP 10 mM ;
- 2 μ l solution [Tris-HCl 0,2 M pH 8,5 ; MgCl₂ 0,1 M] ;
- 14 μ l H₂O ultrapure.

- La solution de bleu de formamide

C'est la solution de dépôt qui arrête la synthèse réalisée durant la chasse. C'est elle aussi qui, par gravité, concentre l'échantillon au fond du puits du gel juste avant la migration. Cette solution possède la composition suivante:

- 100 ml formamide (désionisée à l'Amberlite et filtrée) ;
- 4 ml EDTA 0,5 M ;
- 0,1 g xylène cyanol ;
- 0,1 g bleu de bromophénol.

Remarque : Dans les cas où le séquençage a porté sur des régions de DNA riches en G et C, nous avons utilisé préférentiellement le 7-Déaza-déoxyGTP à la place du dGTP. Cette modification a permis d'éviter les problèmes de compression des bandes sur les gels lors du séquençage de segments riches en G et C. La lecture est alors facilitée et les risques d'ambiguïtés sont amoindris.

1-3-2. Hybridation de l'amorce sur le DNA sb

Cette étape est réalisée dans les conditions suivantes :

- 5 μ l DNA sb à 0,2 μ g/ μ l ;
- 2 μ l amorce à 2,5 ng/ μ l ;
- 1,5 μ l tampon Klenow 10x [Tris 0,1 M (pH8,0) ; MgCl₂ 0,05 M] ;
- 2,5 μ l H₂O ultrapure.

Le mélange est incubé 5 mn dans un bain-marie thermostaté à 56°C, puis laissé refroidir lentement à température ambiante.

1-3-3. Elongation

Cette étape consiste en la synthèse dans le sens 5' vers 3' du brin complémentaire de DNA à partir de l'amorce.

La solution contenant l' $[\alpha\text{-}^{32}\text{P}]\text{-dATP}$ est une solution alcoolique à 50 % (v/v). Elle présente une activité spécifique de 400 Ci/mme. Cette solution est évaporée à sec à raison de 12 μl , soit 12 μCi par séquençage. Le dATP radioactif sec est repris dans 2 μl d'eau ultrapure.

Les 2 μl d' $[\alpha\text{-}^{32}\text{P}]\text{-dATP}$ et deux unités de DNA polymérase I fragment de Klenow sont ajoutés au mélange d'hybridation. Dans les microtubes notés A, C, G, et T, 2,5 μl de milieu d'hybridation sont additionnés à 2 μl du mélange déoxy/didéoxynucléotide correspondant. Les tubes sont placés à 37°C pendant 20 mn.

1-3-4. Chasse et arrêt de la réaction

En fin d'élongation, 2 μl de solution de chasse sont ajoutés à chaque tube. Après homogénéisation du contenu, les tubes sont placés à 37°C pour une incubation supplémentaire de 20 à 25 mn.

La réaction est finalement arrêtée par l'addition de 4 μl de solution de bleu de formamide dans chaque tube. A ce stade, les échantillons peuvent être congelés aussitôt ou traités en vue de leur dépôt sur le gel de séquençage.

2. Adaptation de la technique de Sanger au séquençage sur plasmide

Dans certaines conditions, il peut s'avérer nécessaire de séquencer un fragment de DNA inséré dans un plasmide afin d'éviter son sous-clonage dans le phage M13, s'il existe un risque d'instabilité et de recombinaison dans les séquences à déterminer.

Pour ce faire, nous avons adopté un protocole dérivé de la technique de Sanger. Basé sur le même principe d'arrêts spécifiques des chaînes en élongation par l'insertion de didéoxynucléotides, ce protocole a été optimisé pour le séquençage de plasmides dérivés de pUC. Ceux-ci présentent l'avantage de posséder en aval de leur fragment "polylinker", une séquence complémentaire permettant la fixation spécifique de l'amorce universelle. Cependant, l'utilisation d'oligonucléotides autres que l'amorce universelle a été nécessaire pour le séquençage des parties

internes des fragments insérés dans les vecteurs pUC. Ces amorces ont été synthétisées au laboratoire et utilisées également pour le séquençage de segments sous-clonés dans les plasmides dérivés de pBR322 ; les résultats obtenus à partir de ces derniers ont cependant été de moindre qualité.

Le procédé adapté au séquençage sur plasmides pUC comprend une étape initiale supplémentaire, la dénaturation alcaline, permettant d'obtenir l'ouverture du double brin nécessaire à l'hybridation de l'amorce.

Une quantité de 2 µg de DNA plasmidique est précipitée à l'éthanol. Le DNA est repris dans 40 µl de solution [NaOH 0,2 M ; EDTA 0,2 mM ; pH 8,0] et incubé 5 mn à température ambiante. La solution est alors neutralisée par 4 µl d'une solution d'acétate d'ammonium 2 M à pH 4,5. Le DNA est alors précipité dans deux volumes d'éthanol à -80°C pendant 5 mn. Le culot est séché après centrifugation de 15 mn à 13.000 tr/mn et lavage à l'éthanol 70 % (v/v). A ce stade, le DNA dénaturé peut être conservé quelques jours à -20°C.

L'hybridation de l'amorce est réalisée en reprenant le DNA dénaturé dans le mélange :

Amorce (5 µg).....	1	µl
Tampon Klenow 10x.....	1,5	µl
dATP (12 µCi).....	1	µl
H ₂ O ultrapure.....	qsp	15 µl

L'hybridation est réalisée à 37°C pendant 15 mn.

L'élongation est menée aussitôt après avoir ajouté deux unités de l'enzyme de Klenow et réparti le mélange dans quatre microtubes contenant les mélanges déoxy/didéoxyribonucléotides.

L'élongation, la chasse et l'arrêt sont réalisés comme décrit pour le séquençage sur le phage M13.

3. Séparation électrophorétique

Le gel de séquençage est un gel de polyacrylamide à 6 % (p/v) dénaturant (urée 7 M).

Pour un gel de dimensions 400 mm x 300 mm x 0,5 mm, le mélange suivant est réalisé :

- 15 ml solution acryl/bisacryl (38/2) 40 % (p/v) après désionisation à l'Amberlite ;
- 5 ml tampon TEB 20x ;
- 42 g urée.

L'urée est dissoute à 37°C et le volume est ajusté à 100 ml avec de l'eau ultrapure. Le gel est polymérisé en ajoutant le TEMED et le persulfate d'ammonium dans les conditions déjà décrites.

Le gel est préchauffé pendant 1 h à 1.200 V pour qu'il atteigne la température de 60°C. Les quatre microtubes de chaque réaction de séquençage sont chauffés à 100°C pendant 3 mn. Chaque échantillon est immédiatement déposé à raison de 1 µl dans les puits du gel. La migration est réalisée dans le tampon TEB, en assurant une température constante et uniforme de 60°C environ.

La solution de bleu de formamide, utilisée comme solution de dépôt, contient deux colorants, le bleu de bromophénol et le bleu de xylène-cyanol, comigrant respectivement avec des chaînes d'environ 20 et 130 nucléotides de long. Ils peuvent donc servir également de témoins de migration.

Ainsi, un gel court, arrêté après 2 h 30 de migration environ (lorsque le bleu de bromophénol est au bas du gel), permet de lire une séquence de près de 180 nucléotides à partir d'environ 15 nucléotides en 3' de l'amorce. Un temps de migration de 6 h est nécessaire pour un gel moyen (1 h après la sortie du bleu de xylène-cyanol) et de 10 h pour un gel long. Les gels moyen et long permettront de déterminer la séquence de 200 à 250 nucléotides supplémentaires, soit au total pour les trois gels environ 400 nucléotides.

4. Traitement du gel et autoradiographie

Après migration, le gel est démoulé et traité à l'acide acétique à 10 % (v/v) pendant 10 mn. Ce traitement permet de fixer les fragments de DNA en les précipitant afin de conserver la qualité de haute résolution du gel lors de l'autoradiographie.

Le gel est épongé et transféré par adhérence sur une feuille de papier Whatman 3MM et recouvert d'un film plastique. Le gel est alors séché sous vide à la température de 80°C pendant 1 h. Il est enfin recouvert d'un film Fuji RX et placé en autoradiographie pour une durée variant de 12 à 72 h selon la radioactivité incorporée.

VIII - MUTAGENESE DIRIGEE

1. Principes de la technique (Kramer et al., 1984)

Nous avons utilisé une technique de mutagénèse performante mise au point au laboratoire. Un hétéroduplex est réalisé entre le phage M13mp9 sb dans lequel est inséré le fragment de DNA à muter et le phage M13mp19 db coupé par les enzymes de restriction HindIII et EcoRI (Fig. 22). Au duplex ainsi formé, on ajoute l'oligonucléotide de synthèse phosphorylé en 5' par la polynucléotide kinase du phage T4. La DNA polymérase (fragment de Klenow) achève la synthèse du brin complémentaire en présence des quatre déoxynucléotides. La DNA ligase du phage T4 circularise l'ensemble par ligation.

La formation de l'hétéroduplex présente un double avantage : il réduit la longueur des fragments de DNA à synthétiser par la DNA polymérase, tout en diminuant les risques de fixation aspécifique de l'oligonucléotide sur le phage M13mp9.

La souche HB2154 est ensuite transformée par le mélange de ligation. Cette souche ne possède pas de tRNA suppresseurs des mutations ambre et présente un système de réparation non-fonctionnel. Après transformation, chaque brin de l'hétéroduplex conduit à une forme répliquative double brin différente dans la cellule-hôte. Le double brin provenant du phage M13mp9 et renfermant le gène sauvage ne pourra donner de phages infectieux puisqu'il possède deux codons ambre. Par contre, le double brin provenant du phage M13mp19 renfermant le gène muté donnera des phages infectieux.

2. Etapes

2-1. Formation de l'hétéroduplex

Cette étape consiste à réaliser un hétéroduplex entre le phage M13mp9 sb portant le gène à muter et le phage M13mp19 db coupé par HindIII et EcoRI.

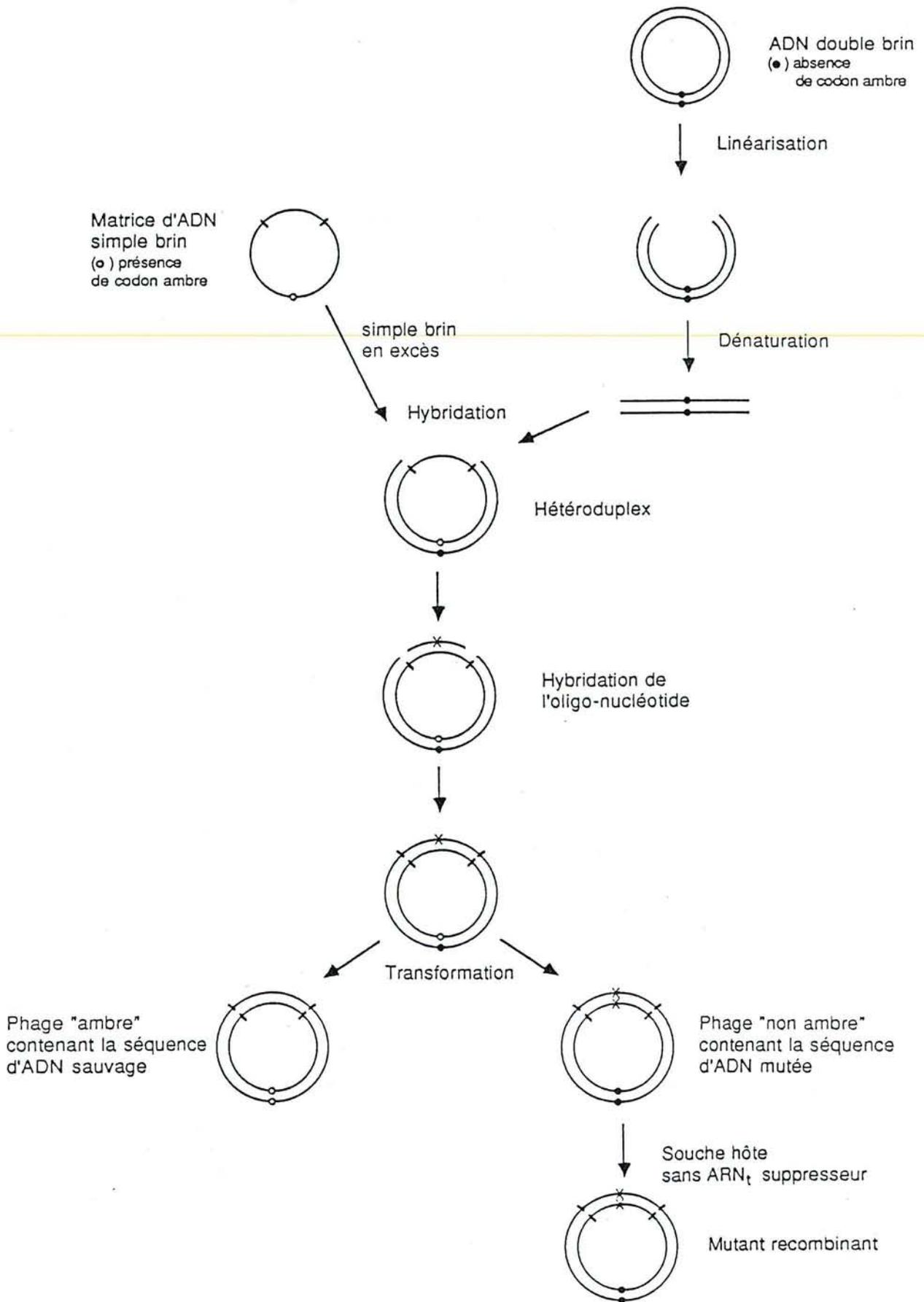


Figure 22 : Principe de la méthode de mutagenèse dirigée utilisée

Le mélange M1 est composé de :

- 1 µg M13mp19 db coupé par HindIII et EcoRI ;
- 2,5 µg M13mp9 sb portant le fragment de DNA à muter ;
- 4 µl tampon Mut 5x [Tris-HCl 62,5 mM ; KCl 937,5 mM] ;
- H₂O ultrapure qsp 20 µl.

Ce mélange est incubé à 100°C pendant 3 mn, à 65°C pendant 5 mn, puis refroidi lentement à température ambiante. La formation de l'hétéroduplex est contrôlée sur gel d'agarose à 0,8 % (p/v).

2-2. Phosphorylation de l'oligonucléotide

Cette étape consiste à placer un groupement phosphate à l'extrémité 5' de l'oligonucléotide. Ce groupement est indispensable à la liaison phosphodiester permettant d'intégrer l'oligonucléotide dans l'hétéroduplex. Cette phosphorylation est réalisée en présence d'ATP duquel le groupement phosphate en γ est transféré sur l'oligo par la polynucléotide kinase du phage T4.

Le mélange suivant est réalisé :

- 1 µg oligo ;
- 1 µl tampon kination 10x [Tris-HCl 0,5 M ; DTT 50 mM ; MgCl₂ 100 mM ;
EDTA 1 mM] ;
- 2 µl ATP 10 mM ;
- 1 µl (4,5 U) T4 polynucléotide kinase ;
- H₂O qsp 10 µl.

Le mélange est incubé 1 h à 37 °C, puis 10 mn à 65 °C. L'oligonucléotide phosphorylé est ensuite traité au phénol et au chloroforme, puis précipité à l'éthanol. Il est repris dans de l'eau et conservé à -20 °C.

2-3. Hybridation de l'oligonucléotide

L'oligonucléotide porteur de la mutation est fixé sur l'hétéroduplex formé dans le mélange M1.

Le mélange M2 est réalisé en fixant un rapport DNA sb/oligonucléotide de 1/20. L'oligonucléotide est utilisé en excès afin de saturer tous les hybrides formés.

L'hybridation est réalisée dans le tampon [Tris-HCl pH 7,5 0,2 M ; MgCl₂ 0,1 M ; NaCl 0,5 M ; DTT 10 mM]. Le mélange M2 est incubé 5 mn à 55°C, puis refroidi lentement à température ambiante.

2-4. Elongation et ligation

Durant cette étape, la DNA polymérase complète l'hétéroduplex, tandis que la DNA ligase le circularise. Pour cela, le mélange M3 suivant est réalisé :

- 1 µl tampon [Tris-HCl (pH 8,5) 0,2 M ; MgCl₂ 0,1 M ; DTT 10 mM] ;
- 1 µl chaque dNTP 10 mM ;
- 1 µl ATP 10 mM ;
- 2,5 U DNA polymérase I (Fragment de Klenow) ;
- 1,5 U DNA ligase du phage T4 ;
- 1 µl DTT 100 mM ;
- H₂O ultrapure qsp 10 µl.

Un même volume de mélange M3 est ajouté au mélange M2. L'ensemble est incubé une nuit à 15°C. Le lendemain, 1 U d'enzyme de Klenow et 1 U de T4-DNA-ligase sont rajoutées pour une incubation supplémentaire de 2 à 3 h à température ambiante.

2-5. Transformation

Une partie du mélange de ligation est utilisée pour transformer les cellules compétentes de la souche HB2154 en utilisant les cellules de la souche HB2151 pour constituer le tapis cellulaire. Ces dernières possèdent un système de réparation fonctionnel capable de corriger les mutations accidentelles pouvant se produire lors de la réplication du phage.

3. Sélection des mutants

A partir des plages obtenues sur milieu LB après transformation de cellules compétentes de HB2154, des préparations de DNA simple brin sont réalisées afin d'identifier les phages porteurs de la mutation.

L'identification est réalisée par séquençage par la méthode de Sanger sur chaque clone étudié.

IX - HYBRIDATION DE DNA PAR LA METHODE DE SOUTHERN

1. Principe (Southern, 1975)

Les possibilités d'hybridation entre différentes molécules complémentaires de DNA constituent une propriété intéressante d'un point de vue analytique. Elles permettent de répondre à des questions telles que un fragment de DNA a-t-il un homologue dans un génome donné ou deux segments de DNA sont-ils contigus dans un génome. Les hybridations peuvent représenter également une stratégie de clonage.

Après séparation par électrophorèse, des fragments de DNA sont transférés sur une membrane de nylon. Une sonde de DNA marquée est hybridée avec le ou les fragments complémentaires qu'elle reconnaît sur la membrane. L'hybride résultant est localisé et identifié par autoradiographie.

2. Marquage d'une sonde de DNA par "random priming"

2-1. Principe (Feinberg et Vogelstein, 1983 et 1984)

Sur un fragment de DNA que l'on veut utiliser comme sonde, on hybride un mélange contenant toutes les séquences d'hexamères possibles. Ceux-ci servent d'amorces à l'enzyme de Klenow pour synthétiser des fragments complémentaires à la sonde. En plaçant dans le milieu réactionnel les quatre déoxynucléotides dont un sous forme marquée, les chaînes incorporent de la radioactivité lors de leur synthèse. Ainsi se constituent des sondes marquées.

Nous avons utilisé pour cela le kit Amersham (Réf. RPN. 1601).

2-2. Etapes

2-2-1. Dénaturation

Un fragment (ou une molécule linéarisée) de DNA est dénaturé afin d'obtenir une forme simple brin. La dénaturation est réalisée dans un microtube dans lequel se trouve un volume de 10 μl contenant de 5 à 25 ng de DNA en solution dans de l'eau ou du tampon TE, placé à 100°C pendant 3 mn, puis immédiatement plongé dans l'eau glacée.

2-2-2. Elongation

Cette étape consiste en la synthèse de chaînes complémentaires à la sonde à partir d'amorces hybridées à celle-ci. Le mélange suivant est constitué :

- 10 μl 5-25 ng DNA dénaturé ;
- 10 μl solution 1 (tampon concentré contenant dATP, dGTP et dTTP) ;
- 5 μl solution 2 (mélange d'hexanucléotides dans l'eau) ;
- H₂O ultrapure qsp 50 μl ;

- 2,5 μl (25 μCi) [α -³²P] dCTP (6.000 Ci/mmol.) ;
- 2 μl enzyme de Klenow à 1 U/ μl dans tampon TE.

L'ensemble est homogénéisé délicatement à la pipette et incubé à 37°C pendant 30 mn. Le DNA marqué est alors congelé à -20°C et conservé ainsi.

2-3. Purification de la sonde marquée

La solution contenant la sonde utilisée pour l'hybridation doit être purifiée afin d'éliminer les nucléotides marqués non incorporés susceptibles de générer un "bruit de fond" lors de l'autoradiographie.

La purification est réalisée par gel-filtration sur résine Sephadex G-25 coulée en micro-colonne dans une pointe jaune de pipette automatique de 200 μl . La sonde, non retenue sur la résine, est éluée en premier, elle est concentrée par précipitation à l'éthanol, reprise dans 20 μl d'eau et conservée à -20°C.

3. Transfert de DNA

3-1. Electrophorèse et traitement du gel

Les produits de digestion et les différents témoins sont déposés sur un gel d'agarose à 0,8 % (p/v). L'électrophorèse terminée, le gel est immédiatement traité de manière à dénaturer le DNA fractionné. Le gel est plongé dans une solution de HCl 0,2 N pendant 15 mn, puis dans une solution [NaOH 0,5 M ; NaCl 1,5 M] pendant 20 mn et enfin dans une solution [Tris 0,5 M ; NaCl 3M] durant 30 mn.

3-2. Transfert sur membrane de nylon

Le transfert des molécules de DNA a été réalisé du gel sur une membrane de nylon (Hybond-N Amersham) par capillarité. Le montage suivant est constitué :

- 2 feuilles de papier Whatman 3MM trempant dans deux bacs contenant SSC 20x [NaCl 3 M ; Citrate-Na 0,3 M ; pH 7,2] ;
- 1 feuille de plastique avec une fenêtre aux dimensions de la partie du gel à transférer ;
- le gel traité ;
- la membrane de nylon ;
- 6 feuilles de papier Whatman 3MM sèches ;
- plusieurs épaisseurs de papier filtre sec ;
- une charge (1 kg environ) répartie uniformément.

Le transfert est réalisé ainsi à température ambiante pendant une quinzaine d'heures.

3-3. Traitement de la membrane

La membrane de nylon est récupérée et lavée dans du tampon SSC 2x. Elle est alors séchée à température ambiante et placée entre deux feuilles de plastique. L'ensemble est posé pendant 3 mn sur une table lumineuse à U.V. (longueur d'onde 310 nm), le DNA étant placé au-dessous (côté U.V.). Ce traitement permet de fixer les molécules de DNA sur la membrane.

4. Hybridation de la sonde

4-1. Préhybridation

Cette étape consiste en une préincubation de la membrane de nylon dans les conditions d'hybridation. Elle permet de diminuer le "bruit de fond" tout en favorisant l'hybridation spécifique de la sonde. La membrane est placée dans un sac en plastique scellé contenant de la solution [SSC 2x ; Denhardt 5x ; SDS 0,1 % (p/v) ; DNA de thymus de veau (soniqué et dénaturé) à 20 ng/ml]. La solution de Denhardt 60x correspond au mélange [Ficoll 1,2 % (v/v) pH 7,0 ; Polyvinylpyrrolidone 1,2 % (v/v)]. L'ensemble est placé pendant 2 h à 65°C.

4-2. Hybridation

La membrane est récupérée et placée dans un autre sac en plastique scellé, dans lequel se trouve la même solution que celle utilisée lors de la préhybridation, mais à laquelle ont été ajoutés environ 10 ng de la sonde marquée après dénaturation par chauffage à 100°C pendant 3 mn et refroidissement rapide dans la glace fondante. L'ensemble est placé à 65°C pour une incubation de 15 h environ.

4-3. Lavages

L'hybridation terminée, le mélange réactionnel est prélevé et conservé à -20°C. La membrane est récupérée et lavée pour éliminer les traces de radioactivité non fixée et responsable du "bruit de fond".

Les lavages sont réalisés à 65°C selon le protocole :

- 2 x 15 mn dans SSC 2x ;
- 2 x 30 mn dans SSC 2x ; SDS 0,1 % (p/v) ;
- 1 x 10 mn dans SSC 0,2x.

La membrane est ensuite séchée à l'air libre.

4-4. Révélation par fluorographie

Un film Fuji RX est placé entre la membrane et un écran intensificateur, le tout entre deux plaques de verre. Les rayonnements β produits par la sonde hybridée sur la membrane de nylon provoquent l'émission de photons au niveau de l'écran. Le signal est ainsi intensifié au niveau du film, d'où l'intérêt de la technique de fluorographie dans le cas d'une hybridation faible ou d'une sonde peu marquée. L'ensemble est placé dans une boîte noire à -80°C pendant au moins 15 h.

Après révélation du film, selon la spécificité de l'hybridation et l'importance du "bruit de fond", il peut être nécessaire de procéder à d'autres séries de lavages. Ceux-ci seront réalisés dans les mêmes solutions, mais à des températures croissantes s'approchant de la température de fusion de l'hybride, pour ne conserver que les hybrides les plus stables résultant de la fixation spécifique de la sonde.

X - TESTS D'EXPRESSION DU GENE DE LA GAPDH

1. Préparation des échantillons

1-1. Culture cellulaire

La souche DF221 transformée par les plasmides recombinants est cultivée dans 5 ml de milieu M63 glucose-ampicilline. Après 36 h de culture à 37°C sous agitation, les cellules sont récupérées par centrifugation, lavées dans le tampon A et conservées sous forme de culot à -20°C .

1-2. Préparation de l'extrait brut

Le culot cellulaire est resuspendu dans 1,5 ml de tampon [Tris-HCl (pH 7,0) 20 mM ; EDTA 2 mM]. Les cellules, maintenues dans la glace, sont alors lysées par traitement aux ultrasons (6 x 20 s avec un arrêt de 15 s entre deux sonications).

L'extrait brut est constitué par le surnageant obtenu après centrifugation à 13.000 tr/mn pendant 45 mn. Il est prélevé et conservé à 4°C .

2. Estimation de la quantité d'enzyme

Cette opération est réalisée par électrophorèse en gel de polyacrylamide dénaturant selon la technique de Laemmli (1970). Elle vise à séparer les différentes protéines présentes dans l'extrait acellulaire et visualiser la bande correspondant à la GAPDH. En fonction de l'intensité de celle-ci, il sera possible de déterminer la proportion de l'enzyme parmi les protéines solubles de l'extrait et d'apprécier le taux d'expression du gène correspondant.

2-1. Composition du gel

Le gel utilisé est de dimensions 120 mm x 100 mm x 1mm.

Gel de concentration à 5 % (p/v)

Tris-HCl 0,5 M pH 6,8.....	480	µl
SDS 10 % (p/v).....	50	µl
Acryl/bisacryl (38/2) 40 % (p/v).....	625	µl
TEMED.....	50	µl
Persulfate d'ammonium 10 mg/ml.....	50	µl
H ₂ O ultrapure.....	qsp 5	ml

Gel de séparation à 10 % (p/v)

Tris-HCl 1 M pH 8,8.....	3,125	ml
SDS 10 % (p/v).....	0,125	ml
Acryl/bisacryl (38/2) 40 % (p/v).....	3,125	ml
TEMED.....	0,062	ml
Persulfate d'ammonum 10 mg/ml.....	0,125	ml
H ₂ O ultrapure.....	qsp 12,5	ml

Tampon d'électrophorèse

Glycine.....	14,4	g
Tris.....	3,0	g
SDS.....	1,0	g
H ₂ O ultrapure.....	qsp 1	l

Le gel est réalisé en coulant tout d'abord le gel de séparation. Du butanol saturé est aussitôt déposé à la surface de celui-ci afin de permettre la réalisation d'un front bien horizontal. Après polymérisation, le butanol est éliminé par plusieurs rinçages à l'eau. Le gel de concentration est alors coulé et un peigne est incorporé à celui-ci.

2-2. Electrophorèse

Les échantillons à déposer sur le gel sont chauffés à 100°C pendant 3 mn après avoir ajouté 1/5^e à 1/10^e de volume d'une solution de dépôt de bleu de bromophénol (utilisée à ce même effet pour l'électrophorèse en gel d'agarose).

Le volume de surnageant de sonication à déposer doit correspondre à environ 60 µg de protéines. La concentration protéique est évaluée à partir de la densité optique de l'extrait acellulaire mesurée à la longueur d'onde de 280 nm. La migration est réalisée à 30 mA pendant environ 2 h, le temps nécessaire à la sortie du front de migration au bas du gel. La séparation s'effectue selon la taille des protéines et non en fonction de leurs charges, neutralisées par le SDS présent dans le gel.

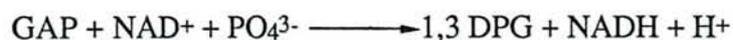
2-3. Révélation et étude des profils

Après migration, le gel est démoulé puis coloré 1 h sous agitation douce dans une solution de bleu de Coomassie à 0,2 % (p/v) dans l'acide acétique à 10 % (v/v). Il est ensuite décoloré dans des bains successifs d'acide acétique à 10 % (v/v). Il est finalement transféré sur papier Whatman 3MM ou placé dans un sac en plastique étanche.

Le gel ainsi conditionné peut alors être utilisé pour l'étude densitométrique du contenu protéique des fractions testées. La lecture du gel est réalisée par l'appareil Ultrosan LKB. L'enregistrement et l'analyse des profils sont effectués par le logiciel Nelson sur ordinateur IBMPC.

3. Mesure de l'activité enzymatique de la GAPDH

La GAPDH catalyse la réaction suivante :



L'activité enzymatique est suivie selon le test de Ferdinand (1964). Il mesure l'apparition du NADH en fonction du temps.

3-1. Solutions nécessaires

3-1-1. Tampon Ferdinand

K ₂ HPO ₄ , 3 H ₂ O.....	50	mM
EDTA.....	0,2	mM
Triéthanolamine.....	40	mM

Le pH est ajusté à 9,2 avec l'acide orthophosphorique.

3-1-2. Préparation du GAP

La résine DOWEX 50 W activée est ajoutée à raison de 1,5 g dans 5 ml d'eau contenant 200 mg de GAP (D.L-glycéraldéhyde-3-phosphate diéthylacetal, monobarium salt ; Sigma Chemical). L'ensemble est chauffé 3 à 4 mn à 100°C, puis filtré ou centrifugé pour éliminer la résine. Celle-ci est lavée avec 4 ml d'eau. La solution finale est ajustée à 10 ml.

Cette étape consiste en une hydrolyse permettant d'obtenir 400 µmoles de mélange de D- et L-glycéraldéhyde-3-phosphate, soit 200 µmoles de D-glycéraldéhyde-3-phosphate, substrat reconnu par la GAPDH. La concentration finale de D-GAP est de 20 mM.

3-2. Test enzymatique

Le milieu réactionnel est constitué ainsi :

D-GAP 20 mM.....	50	µl
NAD ⁺ 100 mM (en solution aqueuse).....	10	µl
Extrait protéique + tampon Ferdinand.....	qsp	1 ml

L'incubation est réalisée en cuve spectrophotométrique dans laquelle l'apparition du NADH est suivie à 340 nm avec un coefficient d'extinction molaire de $6,22 \times 10^3 \text{ M}^{-1} \cdot \text{cm}^{-1}$.

L'activité spécifique de la GAPDH est calculée d'après la formule :

$$\text{A.S.} = \frac{\text{pente} \times \text{dilution} \times 10^3}{\text{volume d'essai} \times 6,22 \times \text{DO}_{280}/\text{ml}}$$

L'activité spécifique est mesurée en $\mu\text{moles NADH} / \text{mn} / \text{mg protéines}$.

RESULTATS ET DISCUSSIONS

SOMMAIRE GENERAL

- Chapitre I : Clonage et séquençage d'un fragment de DNA
génomique de *Clostridium pasteurianum*
contenant le gène de la GAPDH.....page 102
- Chapitre II : Etude des gènes contenus dans le fragment
génomique de *C.pasteurianum* cloné dans pCGap1.....page 124
- Chapitre III : Etude de l'expression du gène de la GAPDH de
C.pasteurianum dans *E.coli*.....page 143
- Chapitre IV : Séquençage et étude d'un fragment de DNA
génomique de *Bacillus stearothermophilus*
contenant le gène de la GAPDH.....page 158
- Chapitre V : Etude phylogénétique de la GAPDH.....page 191

Chapitre I :

Clonage et séquençage d'un
fragment de DNA génomique de
Clostridium pasteurianum
contenant le gène de la GAPDH

SOMMAIRE

I - CONSTRUCTION DE LA BANQUE GENOMIQUE DE <i>C.pasteurianum</i> DANS <i>E.coli</i> BJ 5183.....	page 104
1. Préparation des fragments de DNA génomique de <i>C.pasteurianum</i>	page 104
2. Préparation du vecteur pBR 322.....	page 106
3. Construction de la banque de plasmides.....	page 106
4. Caractéristiques de la banque obtenue.....	page 106
 II - CLONAGE DU GENE DE LA GAPDH DE <i>C.pasteurianum</i> DANS <i>E.coli</i>.....	 page 107
1. Principe de la complémentation.....	page 108
2. Sélection des clones.....	page 108
3. Vérifications du clone obtenu.....	page 108
 III - ANALYSE ET ESSAIS DE REDUCTION DE LA TAILLE DU PLASMIDE pCGap1.....	 page 109
1. Carte de restriction.....	page 109
2. Essais de réduction de la taille de pCGap1.....	page 109
2-1. Elimination du fragment HindIII de 1,8 kb.....	page 109
2-2. Elimination du fragment EcoRI de 2,8 kb.....	page 112
2-3. Sous-clonage du fragment DraI de 1,8 kb.....	page 112
 IV - DETERMINATION DE LA SEQUENCE NUCLEOTIDIQUE DU FRAGMENT DE 3,7 kb CLONE DANS pCGap1.....	 page 115
1. Sous-clonage de petits fragments.....	page 117
2. Sous-clonage de grands fragments.....	page 117
3. Instabilité des fragments sous-clonés.....	page 119
4. Séquençage sur les plasmides.....	page 119
5. Caractéristiques générales de la séquence nucléotidique du fragment cloné..	page 120

I - CONSTRUCTION DE LA BANQUE GENOMIQUE DE *C.pasteurianum* DANS *E.coli* BJ 5183

La stratégie que nous avons utilisée pour construire la banque génomique de *Clostridium pasteurianum* consistait à hydrolyser partiellement le DNA génomique par Sau3A et insérer les fragments obtenus dans le site BamHI du vecteur pBR322, les clones obtenus après transformation d'*E.coli* par les plasmides recombinants constituant la banque.

1. Préparation des fragments de DNA génomique de *C.pasteurianum*

Pour préparer le DNA génomique de *C.pasteurianum*, nous avons été obligés de procéder à la mise au point des conditions d'extraction. C'est ainsi que nous avons mis en évidence la nécessité d'opérer sur des cellules récoltées en début de phase exponentielle de croissance. A ce stade en effet, l'activité DNasique caractéristique des Clostridia était très faible, ce qui permettait de préparer un DNA génomique de bonne qualité.

D'autre part, il nous a fallu établir un protocole de lyse permettant d'obtenir un rendement d'extraction satisfaisant. Ce protocole est décrit dans la partie "Matériels et Méthodes". La grande résistance des parois des cellules de *C.pasteurianum* a nécessité des conditions de lyse particulièrement fortes. Dans les conditions décrites, le rendement de l'extraction était d'environ 150 µg de DNA/g de cellules avec un rapport DO_{260}/DO_{280} de 1,8.

Nous avons alors réalisé des cinétiques d'hydrolyse par Sau3A, visant à déterminer les conditions nécessaires pour obtenir des fragments de DNA suffisamment grands pour contenir l'intégralité du gène de la GAPDH, sans l'être trop cependant afin qu'ils puissent s'insérer aisément dans pBR322. Les différents hydrolysats ont été déposés sur gel d'agarose. Après analyse des profils de fractionnement (Fig. 23), nous avons retenu les conditions suivantes : une hydrolyse de 10 mn à température ambiante avec un rapport enzyme/DNA de 0,2 U Sau3A/µg DNA génomique, conditions permettant de générer des fragments dont la taille était comprise entre 3 et 12 kb.

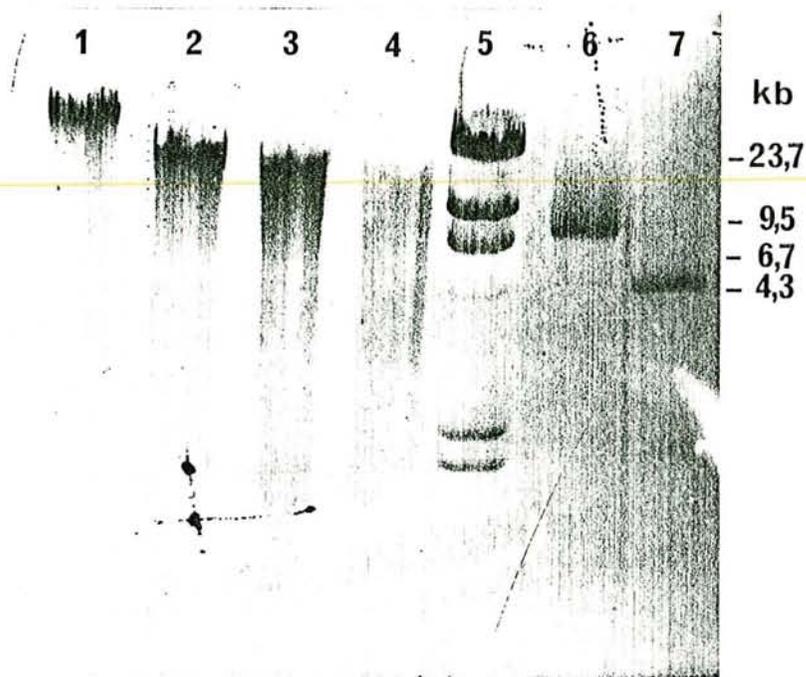


Figure 23 : Cinétiques d'hydrolyse du DNA génomique de *C.pasteurianum*

(gel à 0,8 % d'agarose, visualisation du BrEt sous U.V.)

- 1 : DNA génomique de *C.pasteurianum* non hydrolysé
- 2 : DNA génomique de *C.pasteurianum* après 5 mn d'hydrolyse par Sau3A (0,2 U/g) à température ambiante
- 3 : DNA génomique de *C.pasteurianum* après 10 mn d'hydrolyse par Sau3A (0,2 U/g) à température ambiante
- 4 : DNA génomique de *C.pasteurianum* après 20 mn d'hydrolyse par Sau3A (0,2 U/g) à température ambiante
- 5 : DNA du phage lambda hydrolysé par HindIII
- 6 : Plasmides de la banque
- 7 : Plasmide pBR322 linéarisé par BamHI

2. Préparation du vecteur pBR322

Un vecteur pBR322 a été préparé par hydrolyse par BamHI et déphosphorylation des extrémités 5'. En testant le vecteur par ligation sur lui-même, nous avons pu estimer à environ 5 % la proportion de plasmides qui se recircularisaient sans avoir inséré de fragment de DNA étranger.

3. Construction de la banque de plasmides

L'intégration des fragments Sau3A du DNA génomique dans le vecteur ouvert au site BamHI a été effectuée en utilisant un excès de fragments par rapport au vecteur. La taille de ce dernier était de 4,3 kb tandis que la taille moyenne des fragments génomiques était de 7 kb. Nous avons choisi un rapport pondéral vecteur/fragment de 1/3, ce rapport s'étant expérimentalement avéré être celui permettant d'obtenir un nombre maximal de plasmides recombinés.

Ainsi, 10 µg de vecteur et 30 µg de fragments génomiques ont été soumis à une ligation de 4 h à température ambiante. 7 ml de cellules compétentes d'*E.coli* BJ 5183 ont été transformés par ce mélange de ligation à raison de 150 µl de cellules pour 200 ng de vecteur. Ces 7 ml ont été répartis sur 45 boîtes de milieu LB ampicilline.

Après une nuit d'incubation à 37°C, nous avons dénombré environ 500 colonies par boîte, soit un total de 22.500 clones. Ces derniers ont été récupérés dans 40 ml de milieu LB ampicilline contenant 15 % (v/v) de glycérol et conservés par fractions de 1 ml à 80°C. La solution obtenue présentait une densité optique à 700 nm de 350 unités/ml. Sachant qu'une unité DO équivaut à une concentration de 10^9 bactéries/ml, nous avons obtenu 35×10^{10} bactéries/ml, soit $35 \times 10^{10} / 2,25 \times 10^4 = 1,6 \times 10^7$ bactéries par clone.

4. Caractéristiques de la banque obtenue

Parmi les 22.500 clones obtenus, il nous fallait déterminer le nombre de clones correspondant à des vecteurs refermés sur eux-mêmes sans insertion de fragment de DNA étranger. Pour ce faire, nous avons utilisé la sélection sur tétracycline, l'insertion d'un fragment au site BamHI de pBR322 inactivant le gène de résistance à la tétracycline, tout en conservant la résistance à l'ampicilline.

Un volume de 75 µl d'une dilution 10^{-7} des cellules de la banque mises en stock a été étalé sur milieu LB ampicilline. Environ 2.000 colonies se sont développées en présence d'ampicilline, 100 seulement en présence de tétracycline. Donc, 5 % des clones possédaient un plasmide natif.

Le nombre réel de clones recombinants était donc d'environ 21.000.

Il était important d'estimer la taille moyenne des plasmides de la banque. Le mélange de plasmides issu d'une culture de la banque a été préparé par la technique du lysat clair, puis fractionné par électrophorèse en gel d'agarose à 0,8 % (p/v). La taille des plasmides a été estimée par comparaison de leur mobilité électrophorétique avec celle des fragments HindIII du DNA du phage λ . Nous avons estimé que la majorité des plasmides hybrides possédait une taille comprise entre 8 et 10 kb, soit une taille des fragments insérés de 4 à 6 kb (Fig. 23).

Ces résultats confirmaient le fait que la banque de plasmides construite devait recouvrir la totalité du DNA génomique de *C.pasteurianum*. En effet, le site Sau3A, qui est un site à quatre nucléotides, présente une probabilité théorique d'existence de $1/4^4$, soit un site tous les 256 nucléotides. Ceci ne se vérifiait bien sûr que dans l'hypothèse d'une composition en bases parfaitement statistique. Si l'on estime la taille du génome de *C.pasteurianum* à 4.10^6 pb et si sa séquence était statistique, il devrait exister $4.10^6 / 256 = 15.600$ sites Sau3A dans ce génome. Etant donné que ce génome présente un pourcentage G + C relativement faible (26 à 28 %), on peut estimer que le nombre de sites Sau3A est beaucoup plus faible, environ 10.000. En considérant que les coupures se soient produites d'une manière statistique et en tenant compte de la taille moyenne des fragments insérés, nous estimons que les 21.000 clones obtenus sont suffisants pour recouvrir la banque.

D'ailleurs, en estimant à 4 kb la taille moyenne des fragments insérés dans les plasmides de la banque, l'ensemble des fragments représente $4 \times 21.000 = 84.10^6$ pb, soit plus de 20 fois la taille du génome de *C.pasteurianum*.

II - CLONAGE DU GENE DE LA GAPDH DE *C.pasteurianum* DANS *E.coli*

La banque génomique de *C.pasteurianum* construite dans *E.coli* a été utilisée pour cloner le gène de la GAPDH. La stratégie que nous avons adoptée est la complémentation de la déficience en GAPDH d'une souche d'*E.coli*, DF221 (Hillman et Fraenkel, 1975). Cette technique avait déjà été utilisée pour le clonage des gènes de GAPDH d'*E.coli* et de *B.stearothermophilus* (Branlant et al., 1983).

1. Principe de la complémentation

La souche d'*E.coli* DF 221 possède un gène de GAPDH qui présente une mutation ambre. Ceci confère le phénotype GAPDH- à la souche ; elle ne peut donc pas se développer sur un milieu contenant le glucose comme unique source carbonée, sauf dans le cas où s'est produite une réversion de la mutation. Le taux de réversion est de $1/10^6$. Elle se développe par contre sur un milieu M63-succinate-glycérol, le métabolisme du succinate ne nécessitant pas la GAPDH.

Transformées par un mélange de plasmides, les cellules de DF221 se développent sur milieu M63-succinate-glycérol-ampicilline. Parmi les cellules transformées non révertantes, seules se développeront sur le milieu M63-glucose-ampicilline celles transformées par un plasmide renfermant un gène actif de GAPDH. C'est ainsi que nous pouvions isoler des plasmides portant ce gène.

2. Sélection des clones

Nous avons transformé les cellules compétentes d'*E.coli* DF221 par les plasmides de la banque à raison de 300 ng de mélange de plasmides par boîte. Ceci nous a permis d'obtenir environ un millier de colonies sur une boîte M63-succinate-glycérol ampicilline. Nous avons ainsi préparé 110 boîtes, ce qui a permis de tester la présence du gène de la GAPDH dans environ 110.000 clones. Sur l'ensemble, nous n'avons obtenu qu'une seule colonie qui se développait sur le milieu M63-glucose-ampicilline.

3. Vérifications du clone obtenu

Il nous fallait vérifier que ce clone isolé n'était pas un révertant transformé par un plasmide quelconque.

Une suspension cellulaire de ce clone a été étalée sur milieu LB tétracycline. Aucune colonie n'est apparue, témoignant de l'insertion d'un fragment au niveau du site BamHI de pBR322. Purifié par mini-préparation selon la technique de lyse alcaline, le plasmide du clone a été utilisé pour transformer des cellules compétentes d'*E.coli* DF221. Autant de colonies se sont développées sur milieu M63-succinate-glycérol-ampicilline que sur milieu M63-glucose-ampicilline. Ce résultat démontrait bien que la souche était complétement par le plasmide, lequel portait le gène de la GAPDH de *C.pasteurianum*. Ce plasmide a été appelé pCGap1.

III - ANALYSE ET ESSAIS DE REDUCTION DE LA TAILLE DU PLASMIDE pCGap1

Le plasmide pCGap1 ayant été isolé, il a fait l'objet de nombreuses études visant à déterminer sa taille, puis la réduire au maximum tout en conservant la capacité de compléter la souche DF221. La stratégie adoptée consistait tout d'abord à établir une carte de restriction permettant de situer un certain nombre de sites de coupure utilisés ensuite pour réduire la taille du plasmide.

1. Carte de restriction

Elle a été établie en soumettant le plasmide pCGap1 à des digestions par les enzymes de restriction suivantes : BamHI, DraI, EcoRI, EcoRV, HindIII, NdeI, PstI, SacI, Sall, ScaI et SmaI. Les hydrolysats ont ensuite été fractionnés par électrophorèse en gel d'agarose à 0,8 % (p/v) (Fig.24). Dans ces conditions, nous ne pouvions pas déceler les fragments dont la taille était inférieure à 0,5 kb. Sall, BamHI et PstI linéarisait pCGap1 en une bande d'environ 8 kb. ScaI, EcoRI, HindIII et NdeI libéraient deux fragments. DraI génèrait deux grands fragments et plusieurs plus petits. SmaI et SacI n'ont par contre pas coupé pCGap1.

Connaissant le nombre et la nature des sites de restriction de pBR322, nous avons pu déduire l'existence, dans le fragment génomique inséré dans pCGap1, d'un site unique pour chacune des enzymes EcoRI, HindIII, ScaI et NdeI.

Nous avons ainsi pu estimer la taille du plasmide pCGap1 à environ 8 kb et celle du fragment génomique inséré à environ 3,7 kb. Une carte de restriction préliminaire a ainsi pu être établie (Fig. 25).

2. Essais de réduction de la taille de pCGap1

Les sites localisés à partir de cette carte de restriction ont été utilisés afin de délimiter la région contenant le gène de la GAPDH.

2-1. Elimination du fragment HindIII de 1,8 kb

La digestion du plasmide pCGap1 par HindIII génère deux fragments de 6,2 et 1,8 kb. Nous avons testé si le fragment de 6,2 kb était suffisant pour compléter la souche DF221.

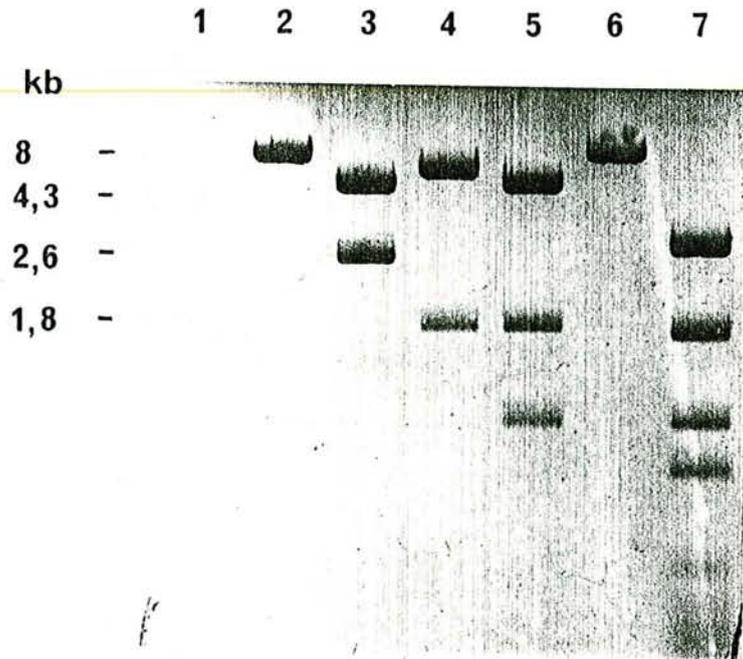


Figure 24 : Fractionnement en gel d'agarose des produits d'hydrolyse de pCGap1

(gel à 0,8 % d'agarose, visualisation du BrEt sous U.V.)

1 : pBR322 / BamHI	5 : pCGap1 / HindIII + EcoRI
2 : pCGap1 / PstI	6 : pCGap1 / BamHI
3 : pCGap1 / EcoRI	7 : pCGap1 / DraI
4 : pCGap1 / HindIII	

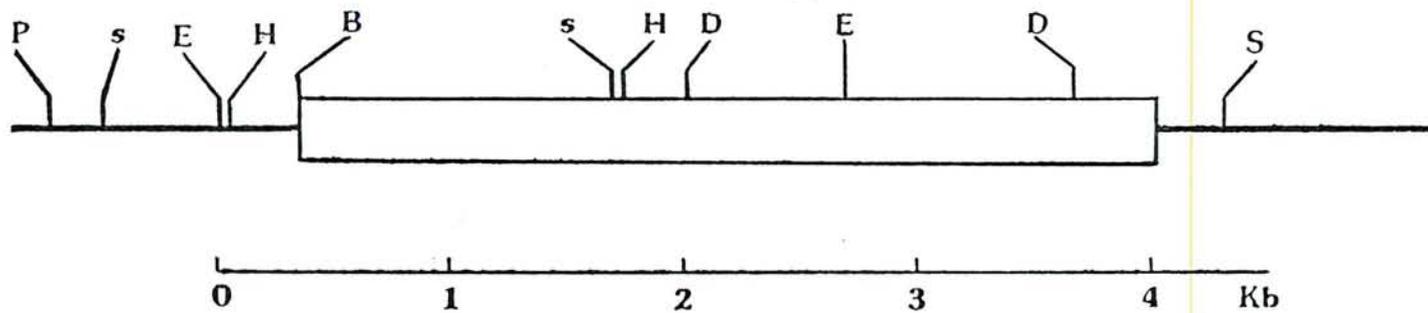


Figure 25 : Cartographie des principaux sites de restriction de pCGap1

B: BamHI - D: DraI - E: EcoRI - H: HindIII - P: PstI - S: SalI - Sca: ScaI

Le plasmide pCGap1 a donc été hydrolysé par HindIII. Après élimination de l'enzyme par extraction phénolique, l'hydrolysate a été soumis à une ligation réalisée dans des conditions de forte dilution favorisant la ligation intramoléculaire. La transformation de cellules compétentes de la souche HB101 par le mélange de ligation a permis d'obtenir une cinquantaine de clones. A partir de ceux-ci, nous avons pu isoler un plasmide de 6,2 kb, qu'il était possible de linéariser par HindIII. Nous l'avons appelé pCGap2 (Fig. 26 et 27).

Nous avons vérifié que ce plasmide pCGap2 complétait bien la souche DF221. Le fragment HindIII de 6,2 kb contient donc l'intégralité du gène structural de la GAPDH de *C.pasteurianum*.

2-2. Elimination du fragment EcoRI de 2,8 kb

La digestion du plasmide pCGap1 par EcoRI génère deux fragments de 5,2 et 2,8 kb. Nous avons cherché à savoir si le fragment de 5,2 kb (qui contient environ 4 kb de pBR322) complétait la souche DF221.

Après élimination de l'enzyme EcoRI, l'hydrolysate de pCGap1 a été incubé dans des conditions favorisant les ligations intramoléculaires. Après transformation de cellules compétentes de la souche HB101, nous avons obtenu une vingtaine de clones recombinants à partir desquels a pu être isolé le plasmide pCGap3 de 5,2 kb, linéarisable par EcoRI (Fig. 26 et 27). Ce plasmide, utilisé pour transformer la souche DF221, ne complétait pas cette souche. Il semblait donc que le site EcoRI soit situé dans la partie codante ou dans une région affectant l'expression du gène de la GAPDH.

Ceci démontrait que le fragment HindIII - EcoRI de l'insert contenait des éléments essentiels pour l'expression de ce gène.

2-3. Sous-clonage du fragment DraI de 1,8 kb

La digestion du plasmide pCGap1 par DraI génère deux grands fragments de 3,5 et 1,8 kb, ainsi que plusieurs petits fragments. Lors de l'établissement de la carte de restriction, nous avons observé que le fragment de 1,8 kb contenait un site EcoRI à 0,7 kb de l'une de ses extrémités. Il était donc possible que ce fragment puisse contenir l'intégralité du gène structural de la GAPDH. Aussi, le sous-clonage de ce fragment de 1,8 kb a-t-il été réalisé.

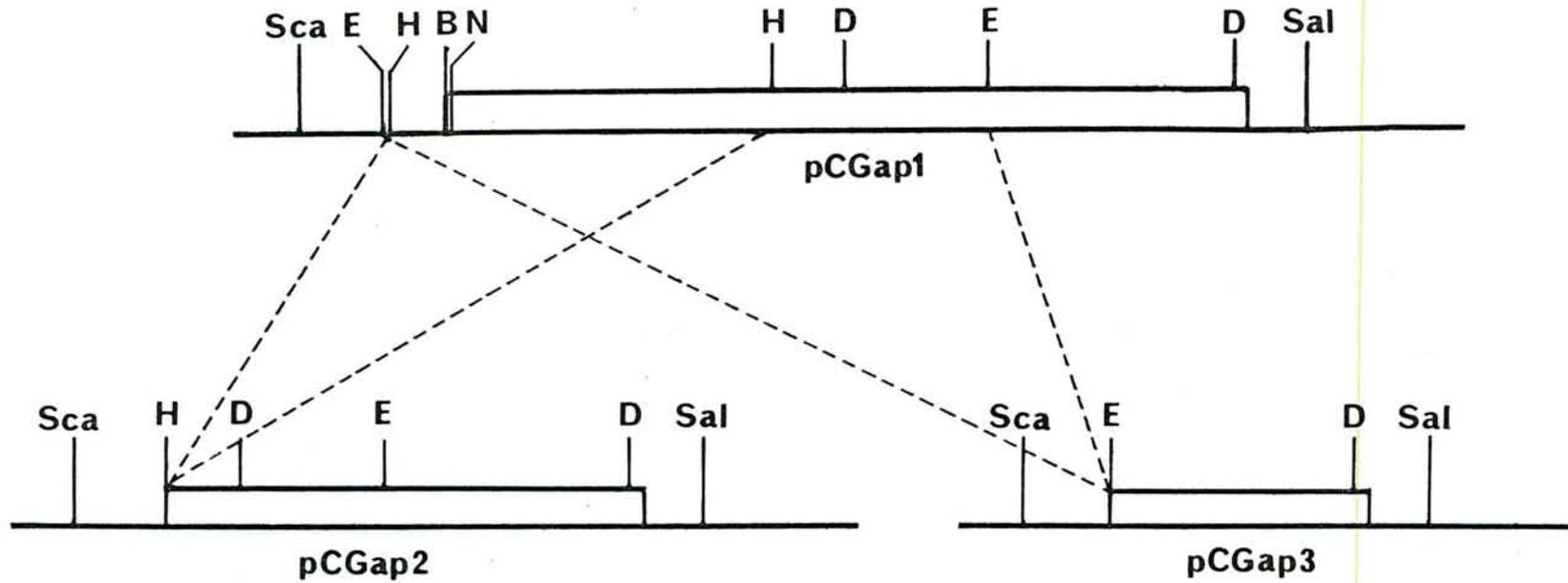


Figure 26 : Construction des plasmides pCGap2 et pCGap3

Le plasmide pCGap2 a été obtenu après élimination du fragment HindIII.

Le plasmide pCGap3 a été obtenu après élimination du fragment EcoRI.

B: BamHI - D: DraI - E: EcoRI - H: HindIII - P: PstI - S: SalI - Sca: ScaI

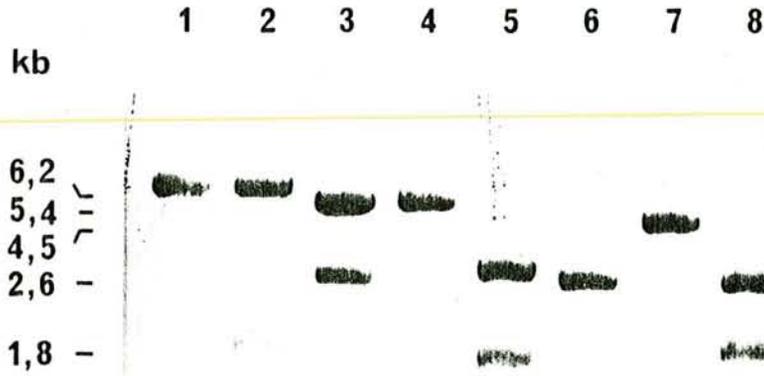


Figure 27 : Fractionnement en gel d'agarose des produits d'hydrolyse de pCGap2 et pCGap3

(gel à 0,8 % d'agarose, visualisation du BrEt sous U.V.)

1 : pCGap2 / HindIII

2 : pCGap1 / HindIII

3 : pCGap1 / EcoRI

4 : pCGap3 / EcoRI

5 : pCGap1 / DraI

6 : pUC19 / HindIII

7 : pCGap4 / BamHI

8 : pCGap4 / BamHI + HindIII

Le fragment DraI de 1,8 kb a été purifié par électrophorèse en gel d'agarose NA à 0,8 % (p/v), électroélué à partir de la bande de gel.

Le fragment purifié a ensuite été inséré dans le site HindII d'un vecteur pUC19, l'incubation ayant été réalisée une nuit à 15°C (Fig. 28). Des cellules compétentes de la souche HB101 ont été transformées par le mélange de ligation. A partir de l'une des colonies obtenues, nous avons préparé le plasmide pCGap4 ayant inséré le fragment DraI de 1,8 kb.

Ce plasmide complète la souche DF221 et possède donc la région codante complète du gène de la GAPDH, ainsi que les éléments nécessaires à son expression.

IV - DETERMINATION DE LA SEQUENCE NUCLEOTIQUE DU FRAGMENT DE 3,7 kb CLONE DANS pCGap1

La détermination de la séquence nucléotidique du fragment de 3,7 kb inséré dans pCGap1 a été réalisée à partir de phages et de plasmides recombinants en utilisant la technique de Sanger et al. (1977).

La stratégie adoptée pour ce séquençage est illustrée par la figure 29.

Les petits fragments ont été obtenus par digestions totales de pCGap1 par des enzymes de restriction reconnaissant des sites à quatre nucléotides, Sau3A, AluI, HaeIII, ou à six nucléotides, DraI. Ces fragments ont été insérés dans des vecteurs M13 appropriés. Ils ont été ensuite séquencés à partir des phages recombinants obtenus.

Les grands fragments BamHI - HindIII et HindIII - EcoRI de 1,4 et 1 kb respectivement ont été sous-clonés dans des vecteurs M13mp9 et M13mp18, l'utilisation de l'amorce universelle a permis de séquencer les extrémités de chacun de ces fragments. L'utilisation d'amorces synthétisées au laboratoire a permis de séquencer leurs régions internes.

Le séquençage sur plasmides s'est avéré indispensable pour achever la détermination de la séquence nucléotidique du fragment complet. En effet, certains sous-fragments insérés dans le phage M13 présentaient une instabilité. C'était le cas du fragment EcoRI - SalI de 1,5 kb et de certains de ses sous-fragments. De ce fait, cette région a été séquencée sur les plasmides pCGap3 et pCGap4 à l'aide de différentes amorces que nous avons synthétisées au laboratoire.

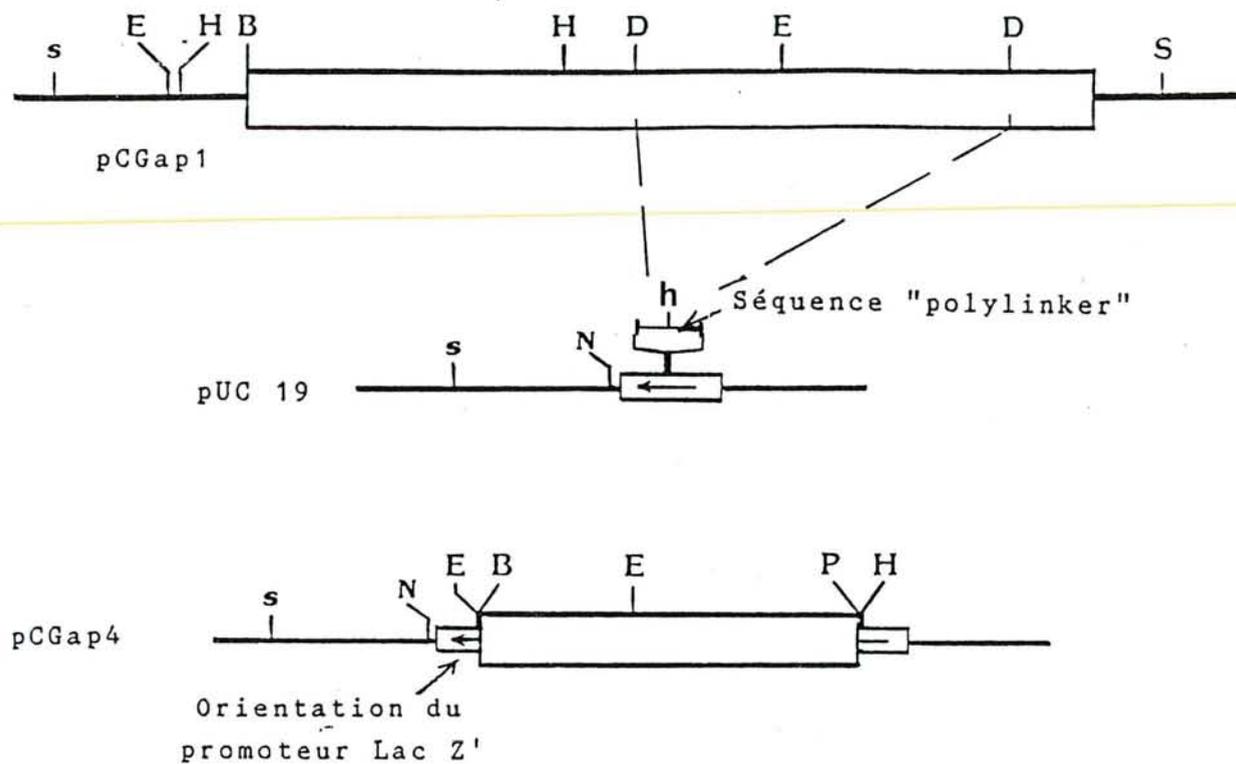


Figure 28 : Sous-clonage du fragment DraI de 1,8 kb

Le fragment DraI de 1,8 kb de pCGap1 est introduit dans le site HindII de la séquence "polylinker" du plasmide pUC19. La construction ainsi obtenue est appelée pCGap4.

B: BamHI - D: DraI - E: EcoRI - h: HindII - H: HindIII - N: NdeI - Sal: SalI

1. Sous-clonage de petits fragments

Après fractionnement en gel de polyacrylamide à 6 % (p/v) des hydrolysats Sau3A, AluI, HaeIII et DraI, les bandes contenant les fragments ont été récupérées. Le DNA a été électroélué et précipité à l'éthanol.

Les fragments à bouts francs AluI, DraI et HaeIII ont été sous-clonés dans le vecteur M13mp18 ouvert par HindIII et phosphatasé.

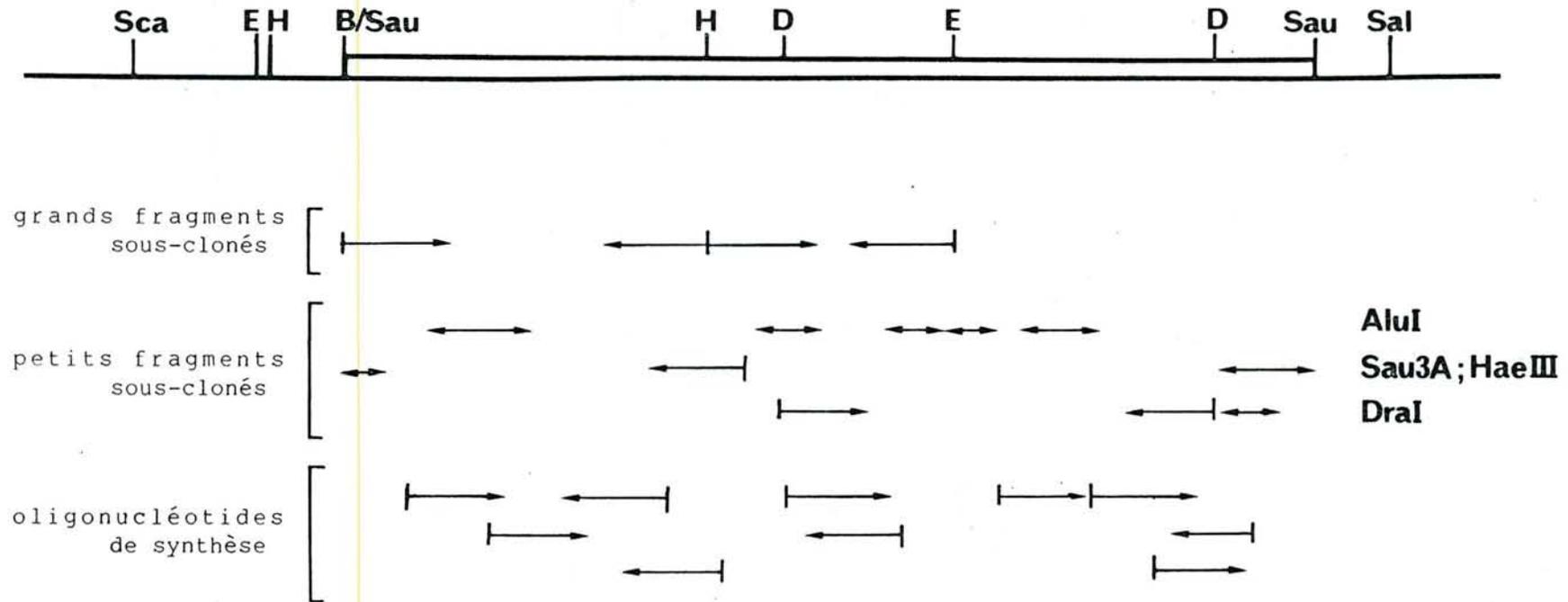
Les mélanges de ligation ont servi à transformer des cellules compétentes d'*E.coli* JM101. Pour chaque boîte, des simples brins ont été préparés à partir d'une dizaine de plages de croissance ralentie. Les phages ayant inséré des fragments dans une orientation opposée ont été sélectionnés par la technique d'hybridation décrite dans "Matériels et Méthodes". Dans le cas où un même fragment avait été inséré dans les deux orientations, la séquence a pu être déterminée dans les deux sens.

2. Sous-clonage de grands fragments

Les fragments BamHI - HindIII et HindIII - EcoRI (Fig. 25) ont été sous-clonés dans les deux orientations en utilisant les vecteurs M13mp9 et M13mp18 coupés par les enzymes correspondantes.

Par hydrolyse totale par AluI, DraI, HaeIII et Sau3A des formes double brin des vecteurs contenant ces grands fragments, il a été possible de classer les fragments libérés par hydrolyse du plasmide pCGap1 dans les trois régions : BamHI - HindIII, HindIII - EcoRI et EcoRI - Sall.

Nous avons pu déterminer la séquence des 350 nucléotides aux extrémités des grands fragments BamHI - HindIII, HindIII - EcoRI à l'aide de l'amorce universelle. Grâce à des oligonucléotides synthétisés à partir des séquences des extrémités déjà connues, le séquençage a pu être prolongé dans les régions internes de ces fragments (Fig. 29).



Sites de restriction : B : BamHI ; D : DraI ; H : HindIII ; E : EcoRI ; Sal : SalI ; Sau : Sau3A ; Sca : ScaI
(le sens de lecture est indiqué par la flèche).

Figure 29 : Stratégie du séquençage de pCGap1

Les fragments sous-clonés dans M13 ont été séquencés à partir de l'amorce universelle. Les régions instables sous forme simple brin ont été séquencées dans les plasmides à l'aide d'oligonucléotides synthétisés au laboratoire.

3. Instabilité des fragments sous-clonés

Certains fragments n'ont pas pu être sous-clonés de manière stable dans le phage M13.

Cette instabilité se traduisait par l'obtention de phages présentant une taille différente de celle attendue, par la perte de sites de restriction, en particulier des sites ayant servi à l'insertion. Cette instabilité était essentiellement dans le cas du fragment EcoRI- SalI de 1,5 kb, mais également de certains de ses sous-fragments, dont la taille était inférieure à 600 pb. Nous avons également observé une instabilité de la région située en 3' du site BamHI.

Graves et Rabinowitz (1986) et Daldal et Applebaum (1985) avaient rencontré ce même problème, lors de l'étude des gènes de la ferredoxine et de la galactokinase de *C.pasteurianum*, certains fragments sous-clonés dans les phages M13 s'étant également avérés instables.

L'instabilité de vecteurs recombinants a déjà été observée dans le cas où le DNA inséré présente un site très fort de démarrage de la transcription. Des instabilités ont également été observées dans le cas de la présence de séquences inversées répétées reconnues par le système de recombinaison de la cellule-hôte. Il nous fallait déterminer la séquence des fragments insérés pour savoir si nous avions à faire avec l'un de ces cas.

4. Séquençage sur les plasmides

Du fait de cette instabilité, le séquençage des régions correspondantes a dû être réalisé sur plasmide, les fragments concernés étant stables dans les plasmides. Le séquençage du fragment EcoRI - SalI a été réalisé sur les plasmides pCGap3 et pCGap4 à l'aide d'amorces synthétiques. Pour cela, nous avons dû mettre au point une variante de la méthode de Sanger, différant par la nécessité d'une dénaturation de la structure double brin du plasmide. Malgré cette mise au point, il s'est avéré que le séquençage sur plasmide était moins performant que sur le phage M13. Il nous permettait de déterminer une séquence d'environ 250 nucléotides à partir de l'amorce. Ceci s'explique sans doute par des arrêts de la progression de la DNA polymérase résultant vraisemblablement de la renaturation de la forme double brin des plasmides lors de la progression de l'enzyme.

5. Caractéristiques générales de la séquence nucléotidique du fragment cloné

La stratégie adoptée nous a permis de déterminer la séquence nucléotidique du fragment complet inséré dans le plasmide pCGap1 (Fig. 30).

Ce fragment comporte 3.663 pb. Chaque séquençage a été répété au moins deux fois. 76 % de la séquence a été définie dans les deux sens et 90 % a été déterminée à partir de deux fragments différents au moins.

Ces précautions nous permettent d'être absolument certains de cette séquence, d'autant plus qu'avec la composition particulièrement riche en A + T du fragment de DNA à séquencer, nous n'avons pas été confrontés au problème de formation de structures secondaires lors du séquençage ou lors du fractionnement sur gel.

La connaissance de la séquence nucléotidique nous a permis de confirmer la carte de restriction préalablement établie et de l'affiner.

Le pourcentage G + C du fragment séquencé est de 28,7 %, valeur voisine de celle rapportée pour *C.pasteurianum*, 26 à 28 %, par Cummins et Johnson (1971).

Comme nous l'avons déjà souligné, certains fragments sous-clonés dans le phage M13 présentaient une forte instabilité lorsqu'ils se trouvaient sous forme simple brin (Fig. 30). Nous avons recherché si des séquences inversées répétées pouvaient expliquer ces remaniements. Aucune structure remarquable de ce type n'a cependant été identifiée. Nous avons seulement remarqué, dans le fragment EcoRI - SalI, la présence d'un octamère GCTGCTGC en position 2.889, présentant 75 % d'homologie avec la séquence chi, connue pour sa faculté de provoquer des recombinaisons chez *E.coli*. La séquence chi semble induire des recombinaisons RecA-dépendantes. Or, les recombinaisons que nous avons observées se produisaient même dans des souches RecA⁻.

Dans l'état actuel de notre étude, il nous est difficile d'expliquer les raisons de cette instabilité, même en ayant déterminé la séquence de la région instable. Ce type d'instabilité a également été décrit dans le cas de plasmides de *Bacillus subtilis*, plasmides qui passent par une phase simple brin lors de leur répllication (Ehrlich, 1988).

Figure 30 : Séquence nucléotidique du fragment de 3.663 pb cloné dans pCGap1 et contenant le gène de la GAPDH de *C.pasteurianum*

```

GATCCATATGTTTTATATTATTGAATCTATAAATAATTTAGGA 44
TATTTATCAGAGAGTATAGAAAATTTATCTAATAGCTTAAAGGTAGATTCTAAATATATAAGTGAAGCTCT 115
AAAAGTAATACAGTCACTAGATCCTACGGGTATTGGAGCTAGAGATTTAAAGGAATGTTTAAATGATACAAG 186
CTAGGAAAAAAGGTATATTAGATAACAACTAAGGAAATAATAGATAGATATTTAGAAGATATAGCTGAT 257
AATAATATAATAAAATTCGCGAAGATATGAATATAACAATTCATGAAGCTCAAATTTATGGAGACATAAT 328
AAAATCCTTTGAACCAAAACCTCTAGGGGATTTTACACAGGTGATGAACTGAATATATAATACCAGATG 399
CTTATATTAAGCAAGTCAAAGAAGATTATATATTTTAAATGAATGACAATTTTATCCCAAGATTCAATT 470
AATAATATATAAAGAATATTTTAAATAATCTCAGGATGAAGATGCTAAAAAATATGTTAAAGAAAAGAT 541
AAATGGTGCTCTTTTTTAAATAAAGTATAGATATGAGAAGAAATACTATTTATAGAGTTATTGAACAAAT 612
TGTTGATATTTCAAAGAAATTTTTTTAGATAAGCATAGTTATTTAAACCATGACTTTAAAGATATTTTC 683
TGAGAGAATTAATATGCATGAATCTACTATCAGCAGAGCTATTAGAGATAAATATATAAATACTGATAGAG 754
GTATTATGAAAATTAAGATTTTATTTACTACTCCATAAGTAATAATAAAAAATGAGGATATATCTACAATA 825
AAAGTGAAAAATATATTAGGGATATAATTGATAAGGAAGATAAATCAAAACCCATTTCCAGATCAGCAAAT 896
AAGTGATATTTTAAAAAATGAAGGCATGAAAATTTCTAGAAGAACTATAGCAAAATATAGGGATGAAATGG 967
AAATAAAATCATCTAGTAAAAGAAAAAGATTTTAGACTTTTTAGTATTTTAAATAAATAATTTAGATTTTT 1038
TTTTAAATACTTATTTAAAAATAAAAAATTTTTATTTATAATATTAAGTGGGACATGAATTGAGTTACTG 1109
GGACATCATAAGGCCAGAGAGGTGGTTAATTCAGGAAGTTTTAAAGTTACAGCAAAAAATGTTCCAGAG 1180
CTT ATG GAC TTA CTT GAA AAA CGA TAC AAT ATT CTT AGA GCA ATT TAC TAT AAT 1224
  H D L L E K R Y N I L R A I Y Y N
CAG CCA ATA GGT AGA AGG ATT TTG GCA AAC CAT TTA GAG ATT GGA GAA AGA ATA 1288
  Q P I G R R I L A N H L E I G E R I
GTC AGA ACG GAG ATT AAT TTT TTA AAA AAG CAA AAT TTG ATA AAT ATA AGT ACT 1342
  V R T E I N F L K K Q N L I N I S T
CCG GGA ATG ACT GTA ACC GCA GAA GGT GAA GAA ATA ATT AAT AAG CTT AAA GAC 1396
  P G H T V T A E G E E I I N K L K D
TTT ATT CAC GAA TTT AAA GGA TTA AGT GAA CAA TCT TTG AAG AAG GTT TTA AAA 1450
  F I H E F K G L S E Q S L K K V L K
ATT AAA GAT GTT ATA ATT GTT CCA GGA GAT GTA GAA GAA GAT CAT ACA GTT ATG 1504
  I K D V I I V P G D V E E D H T V H
AAA GAG CTC GGA AGA ACT GCT GCT CAA TAT GTA AAA AGA ATT ATT AAG GAT AAT 1558
  K E L G R T A A Q Y V K R I I K D N
ACA ATT GTA GCA GTT ACT GGC GGT ACT ACT ATA AAA GAA GTA GTT GAT AGT TTC 1612
  T I V A V T G G T T I K E V V D S F
CCC TGT ATT CCA TAT TTT AAA AAT ACC ACT GTA GTT CCT GCA AGA GGA GCT ATG 1666
  P C I P Y F K N T T V V P A R G A H
GAA AGA ATG GTG GAA ACA GAA GCC AAT ACT CTT GCC GCT AGA TTT GCA AAT AAG 1720
  E R H V E T E A N T L A A R F A N K
CTA GGG GCA AAA TAC AAA CTT TTA AGT GTT CCA GAA AAT TTA AGT GAT TTA GCT 1774
  L G A K Y K L L S V P E N L S D L A
TTT AGC ACA ATT CTC AAT GAG AAG AGC ATT AAA AAT GTC ATA AAT ATG CTT CAC 1828
  F S T I L N E K S I K N V I N H L H
AAT GCA GAT GTT GTC ATA TAT GGA ATT GGA AGA GCT GAT GAA ATG AGT ACA AGA 1882
  N A D V V I Y G I G R A D E H S T R
CGA GGA TTA AGT CAA GAA GAA ATT AAT GTT ATA TTG GAA AAA AAG TCT GTA GGT 1936
  R G L S Q E E I N V I L E K K S V G
GAA GCC TTT GGT TAT TAT TTT GAT AAA AAG GGA AAA GTA GTT CAT TCT ACT CCA 1990
  E A F G Y Y F D K K G K V V H S T P
TCT ATA GGA TTG TCT ATG GAA GAT ATA AAA AAA ATA GAT AAC CTT GTA GCA GTG 2044
  S I G L S H E D I K K I D N L V A V
GCT GGA GGA AAA TCT AAA GCA GAT GCC ATA GTT TCT ACT CAG CTT AGT AAT GAC 2098
  A G G K S K A D A I V S T Q L S N D
AAA AAT GTA CTT ATA ACT GAT GAA GGA ACA GCT ATA GAA ATG TTA AAC ATA ACA 2152
  K N V L I T D E G T A I E H L N I T
AAT ACA AAT TAG TAAACAATTTATTTTATTAAGTATTATTTATGATACTTAGTAAATAATATGTTT 2219
  N D N *

```

TGTTAATATATATAAAGCTAATATTATATATTTACATATAATTTAGAATATTTTATTTTGGAGGTATTAAA 2290
 ATG ACA AAA GTA GCT ATT AAT GGT TTT GGA AGA ATA GGA AGA TTA GCT TTA AGA 2344
 H T K V A I N G F G R I G R L A L R
 1 10
 AGA ATT CTT GAA GTA CCT GGT TTA GAG GTA GTT GCA ATC AAT GAT TTA ACT GAT 2398
 R I L E V P G L E V V A I N D L T D
 20 30
 GCA AAA ATG TTA GCA CAC TTA TTC AAA TAT GAT TCA TCA CAA GGT AGA TTC AAT 2452
 A K H L A H L F K Y D S S Q G R F N
 40 50
 GGT GAA ATT GAA GTT AAA GAA GGA GCT TTC GTT GTT AAC GGA AAA GAA GTT AAA 2506
 G E I E V K E G A F V V N G K E V K
 60 70
 GTT TTC GCA GAA GCA GAT CCT GAA AAA TTA CCA TGG GGA GAA CTA GGA ATT GAT 2560
 V F A E A D P E K L P W G E L G I D
 80 90
 GTT GTT CTT GAA TGT ACT GGA TTC TTC ACT AAA AAA GAA AAA GCT GAA GCT CAT 2614
 V V L E C T G F F T K K E K A E A H
 100
 GTA AGA GCT GGT GCT AAA AAA GTT GTT ATT TCA GCT CCT GCT GGA AAT GAC TTA 2668
 V R A G A K K V V I S A P A G N D L
 110 120 122A
 AAG ACA ATA GTA TTT AAC GTT AAT AAC GAA GAT CTT GAT GGA ACT GAA ACA GTT 2722
 K T I V F N V N N E D L D G T E T V
 130 138A 140
 ATA TCA GGT GCA TCA TGC ACA ACT AAC TGC TTA GCT CCA ATG GCT AAA GTA TTA 2776
 I S G A S C T T N C L A P H A K V L
 150 160
 AAT GAT AAA TTC GGA ATA GAA AAA GGA TTT ATG ACT ACA ATC CAT GCT TAC ACT 2830
 N D K F G I E K G F H T T I H A Y T
 170
 AAT GAC CAA AAC ACA TTA GAC GGC CCA CAC AGA AAA GGT GAT TTC AGA AGA GCT 2884
 N D Q N T L D G P H R K G D F R R A
 180 190 190A 191A
 AGA GCT GCT GCT GTC AGC ATA ATC CCT AAC TCA ACT GGT GCT GCT AAA GCT ATC 2938
 R A A A V S I I P N S T G A A K A I
 200 210
 GCT CAA GTT ATC CCA GAA TTA AAA GGT AAA TTA GAT GGA AAC GCT CAA AGA GTT 2992
 A Q V I P E L K G K L D G N A Q R V
 220 230
 CCA GTT CCA ACT GGT TCA GTA ACT GAA TTA ATT TCA GTT CTT AAG AAA AAT GTT 3046
 P V P T G S V T E L I S V L K K N V
 240 250
 ACA GTT GAA GAA ATC AAT GCT GCT ATG AAA GAA GCT GCA AAT GAA TCT TTA GGA 3100
 T V E E I N A A M K E A A N E S L G
 260 270
 TAC ACT GAA GAT GAA ATA GTT TCA GCT GAC GTT GTT GGA ATA TCT TAT GGT TCA 3154
 Y T E D E I V S A D V V G I S Y G S
 280 290
 TTA TTT GAT GCA ACT TTA ACT AAA ATT GTA GAT GTT GAT GGA TCA CAA TTA GTT 3208
 L F D A T L T K I V D V D G S Q L V
 300 301A 301B 304A
 AAA ACT GTT TCT TGG TAT GAT AAT GAA ATG TCA TAC ACT TCA CAA TTA GTT AGA 3262
 K T V S W Y D N E H S Y T S Q L V R
 310 317A 320
 ACT TTA GAA TAT TTC GCT AAA ATT GCA AAA TAG TCTAAAAGTTAATAATTGAAATTTA 3322
 T L E Y F A K I A K *
 330
 AAAATGTGAACATAGTAAATAGGCTCTGTTTTGTAGTTAAGGCTATGAGACTAGGCCITTTTAAAATAAA 3393

TATGTAATATAGATTCTAGTAGATATAAAACATTTTTTAAATAATTTTATGAGGTGAAATGAT ATG 3462
 H

TCA TAT AAC AAA AAG ACA ATT GAA GAT ATT GAT GTT AAA GGT AAA AGA GTA TTA 3516
 S Y N K K T I E D I D V K G K R V L

GTA AGA TGT GAT TTT AAT GTA CCT TTA AAA GAT GGC GTT ATA ACT GAC GAA AAC 3570
 V R C D F N V P L K D G V I T D E N

AGA CTT AAT GGT GCA CTT CCA ACT ATA AAA TAT TTA ACA GAA AAA GGT GGA AAG 3624
 R L N G A L P T I K Y L T E K G G K

GTT ATC CTT TGC TCA CAT CTT GGA AAA GCT AAA GGA CCA 3663
 V I L C S H L G K A K G P

(La région inscrite en caractères gras à partir du nucléotide 2.347 représente le segment instable lors de son sous-clonage dans le phage M13.)

Chapitre II :

Etude des gènes contenus dans
le fragment génomique de
Clostridium pasteurianum
cloné dans pCGap1

SOMMAIRE

I - CARACTERISATION DES DIFFERENTES PHASES OUVERTES DE LECTURE CONTENUES DANS LE FRAGMENT CLONE.....	page 126
II - ETUDE DES SIGNAUX DE REGULATION.....	page 126
1. Initiation de la transcription.....	page 126
2. Terminaison de la transcription.....	page 133
3. Initiation de la traduction.....	page 135
III - ETUDE DE LA COMPOSITION EN BASES.....	page 137
IV - ETUDE DES PHASES OUVERTES DE LECTURE.....	page 137
1. La phase G4.....	page 137
1-1. Usage des codons.....	page 138
1-2. Structure primaire de la protéine.....	page 138
2. La phase G5.....	page 140
3. La phase G3.....	page 141
V - CONCLUSIONS.....	page 142

I - CARACTERISATION DES DIFFERENTES PHASES OUVERTES DE LECTURE CONTENUES DANS LE FRAGMENT CLONE

L'examen de la séquence du fragment cloné nous a permis de caractériser cinq phases ouvertes de lecture que nous avons appelées G1, G2, G3, G4 et G5 (Fig. 31). Ces phases ouvertes de lecture coderaient respectivement pour des protéines de 90, 111, 326, 334 et 68 amino-acides, la phase G5 étant incomplète.

En raison de la localisation dans le fragment DraI de 1,8 kb complétant le mutant GAPDH-, la phase G4 a pu être identifiée comme étant celle codant pour la GAPDH. Ceci a été confirmé par l'examen de la structure primaire de la protéine codée par cette phase ouverte. La phase G4 se situe entre les nucléotides 2.291 et 3.292 (Fig. 30).

Les trois phases G1, G2 et G3 sont situées en amont du gène de la GAPDH, la phase G5 en aval.

Il était important de définir, parmi ces différentes phases ouvertes de lecture, lesquelles constituaient des régions effectivement codantes. Pour cela, nous nous sommes attachés à rechercher, pour chaque phase ouverte, l'existence de signaux possibles de transcription et de traduction, homologues aux séquences consensus établies pour les bactéries Gram⁺. Nous avons aussi étudié la composition en bases de chaque phase.

II - ETUDE DES SIGNAUX DE REGULATION

1. Initiation de la transcription

Les phases G1 et G2 ne présentaient pas de signaux possibles d'initiation de la transcription, ce qui suggérait qu'elles n'étaient pas codantes. D'ailleurs, la composition en bases de ces courtes phases ouvertes montrait un pourcentage en G + C de 24 %, légèrement inférieur à celui de l'ensemble du génome, alors que la phase G4 identifiée comme codant pour GAPDH présente un pourcentage en G + C est de 35 %. Nous en avons conclu que les phases G1 et G2 ne devaient pas correspondre à des gènes.

Par contre, les phases G3, G4 et G5 étaient toutes précédées de signaux putatifs de démarrage de la transcription.

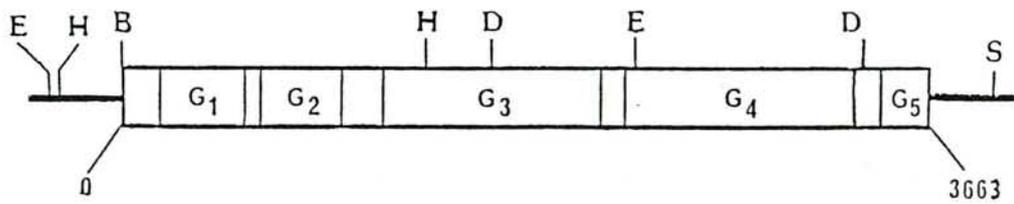


Figure 31 : Les différentes ORF du fragment cloné dans pCGap1

G1 et G2 : non codantes

G3 : codant probablement pour une protéine de 326 aa

G4 : codant pour la GAPDH de C.pasteurianum (334 aa)

G5 : codant probablement pour les 68 aa N-terminaux d'une protéine

B: BamHI - D: DraI - E: EcoRI - H: HindIII - Sal: SalI

Graves et Rabinowitz (1986) ont établi une séquence consensus pour les régions promotrices des gènes de bactéries Gram⁺ : ces régions comportent respectivement les séquences TATAAT (-10) et TTGACA (-35) et, en amont de la séquence -35, une séquence polyA et des segments riches en A et T.

En amont de la phase G3, nous avons observé la présence d'un segment TTGAGT en position -78 et d'un segment TTTATA en position -91, séparés par 18 nucléotides (Fig. 32). Ces deux segments peuvent respectivement correspondre à des boîtes -10 et -35.

En amont de la partie codante du gène de la GAPDH, nous avons identifié deux signaux potentiels d'initiation de la transcription que nous avons appelés a et b. Les séquences de ces deux signaux présentent de fortes homologies avec les séquences promotrices consensus définies pour les gènes d'*E.coli* et de bactéries Gram⁺ (Fig. 33).

Ces séquences sont localisées en -30, TATAAT (séquence -10a), en -43, TTATAT (séquence -10b), en -67, TTGTTA (séquence -35a) et en -82, TTAGTAAA (séquence -35b). Les signaux a et b sont décalés de 7 nucléotides et sont par conséquent chevauchants. Cette organisation particulière, caractéristique des promoteurs de gènes régulés par le cAMP (Spassky et al., 1984), avait déjà été observée au niveau du gène de la GAPDH d'*E.coli* (Branlant et Branlant, 1985), bien que ce gène n'ait pas été décrit comme régulé par le cAMP.

En amont de la phase G5, les séquences TATAAA et TATGTAA, situées respectivement aux positions -36 et -59 par rapport au codon d'initiation et séparées par 17 nucléotides, peuvent correspondre à un promoteur de transcription (Fig. 34).

Les séquences promotrices définies pour chacune des trois régions codantes présentent davantage d'homologies avec les séquences consensus établies pour les régions promotrices des bactéries Gram⁺ qu'avec celles d'*E. coli* (Fig.35). Une étude plus approfondie a été effectuée afin de définir une séquence consensus pour les promoteurs des gènes de *C.pasteurianum* (Assobhei, 1988). Cette séquence consensus présente quelques particularités par rapport à celles établies pour les bactéries Gram⁺ (Fig. 35), à partir de séquences de gènes de bactéries du genre Streptococcus, Bacillus, Corynebacterium et Clostridium, par Graves et Rabinowitz (1986).

ACT TTA GAA TAT TTC GCT AAA ATT GCA AAA TAG TCTTAAAAGTTAATA
 T L E Y F A K I A K *
 ATTGAAATTTTAAAAATGTGAACATAGTAAATAGGCTCGGTTTTGTAGTTAAGGCTATGAGACTAGGCCTTTTTAAAAAC

TAAA TATGTAA +50 ATATAGATT CTAGTAGA TATAAA +30 +20 ATGAGGTGA -1
-35 -10 SD AATGAT

ATG TCA TAT AAC AAA AAG ACA ATT GAA GAT ATT GAT GTT AAA GGT AAA AGA GTA TTA GTA
 M S Y N K K T I E D I D V K G K R V L V
10 20

AGA TGT GAT TTT AAT GTA CCT TTA AAA GAT GGC GTT ATA ACT GAC GAA AAC AGA CTT AAT
 R C D F N V P L K D G V I T D E N R L N
30 40

GGT GCA CTT CCA ACT ATA AAA TAT TTA ACA GAA AAA GGT GGA AAG GTT ATC CTT TGC TCA
 G A L P T I K Y L T E K G G K V I L C S
50 60

CAT CTT GGA AAA GCT AAA GGA CCA
 H L G K A K G P...
68

Figure 34 : Séquence nucléotidique de la phase ouverte G5 et séquence de la protéine correspondante

2. Terminaison de la transcription

Les structures terminatrices de la transcription des gènes des bactéries Gram⁺ séquencés jusqu'à présent sont de même type que celles des gènes d'*E.coli*. Elles se caractérisent par une séquence palindromique riche en G + C, pouvant conduire à la formation au niveau du mRNA d'une structure en "épingle à cheveux" suivie d'un segment polyT.

En aval du codon de terminaison de la phase G3, une séquence palindromique a été observée en position +22 (Fig. 32). L'énergie libre de formation de cette structure est faible, -4,7 kcal/mole (calculs effectués d'après Freier et al., 1986). Cette séquence est suivie d'un segment polyT (Fig. 36).

En aval de la partie codante du gène de la GAPDH, nous avons localisé en position +3 une séquence palindromique riche en A et T, susceptible de constituer une structure secondaire au niveau du mRNA (Fig. 33). L'appariement comprend 8 paires de bases avec une boucle de 8 nucléotides (Fig. 36). L'énergie libre de formation de cette structure est très faible (-0,5 kcal/mole). Cette région palindromique n'est pas suivie d'une séquence polyT. Il est peu probable qu'elle constitue réellement un site de terminaison de la transcription, d'autant plus que cette structure est située à une distance très faible du codon de terminaison.

La phase G5 n'étant pas complète, nous ne pouvons bien sûr rien dire à propos des signaux de terminaison de la transcription de cette phase.

Nous pouvons tirer les conclusions suivantes de cette recherche de signaux d'initiation et de terminaison de la transcription.

- Les phases G3, G4 et G5 sont précédées par des séquences susceptibles d'être des signaux d'initiation de la transcription. Il doit donc bien s'agir de phases codantes.
- La phase G3 est suivie d'un signal possible d'arrêt de la transcription qui, d'après sa structure, est vraisemblablement peu efficace, mais peut-être que le promoteur correspondant a lui aussi une faible efficacité.
- Les phases G4 et G5 ne sont pas séparées par un signal caractéristique d'arrêt de la transcription et doivent donc appartenir à un même opéron, présentant des signaux internes de réinitiation de la transcription.

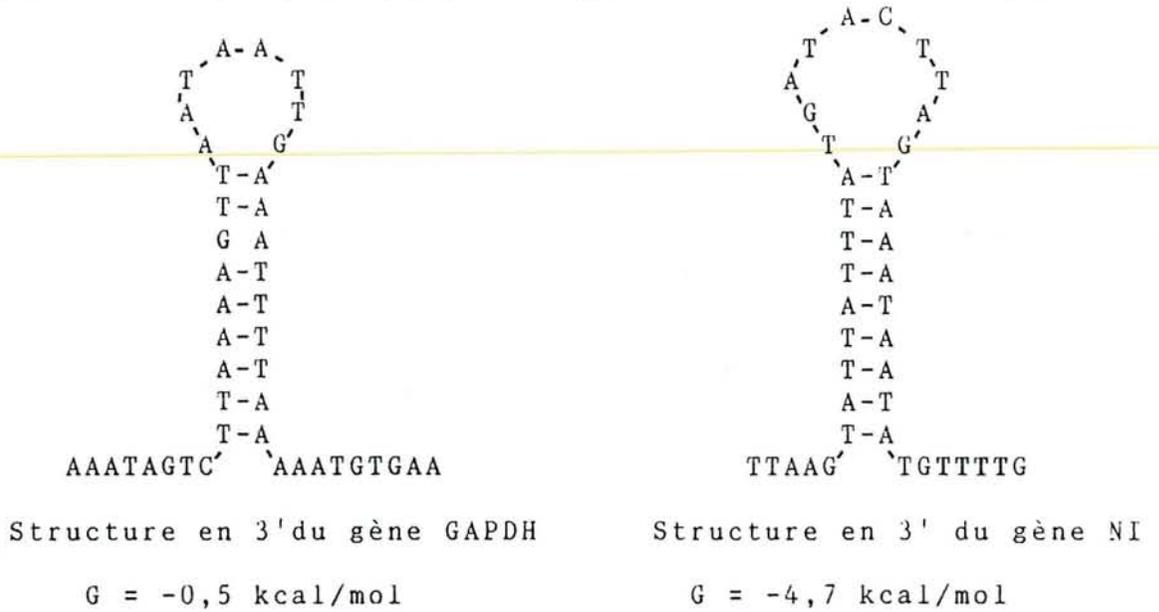


Figure 36 : Signaux potentiels de terminaison de la transcription en aval des gènes de *C. pasteurianum*

3. Initiation de la traduction

Les séquences de reconnaissance du ribosome sur les messagers des bactéries Gram⁺ présentent une forte complémentarité avec l'extrémité 3' du rRNA 16S de ces bactéries (Fig. 37).

Nous n'avons pu caractériser aucune séquence susceptible de correspondre à un site de fixation du ribosome immédiatement en amont de la phase G3. Par contre, deux séquences situées respectivement à 33 et 50 nucléotides en amont du codon ATG (Fig. 32) présentent une grande complémentarité avec l'extrémité 3' très conservée du rRNA 16S des bactéries Gram⁺. La première séquence en -33, AGGAAGT posséderait une énergie libre de liaison avec le rRNA 16S de -5,2 kcal/mole. Celle en -50, AGAGAGGTG, présenterait une énergie libre de fixation de -7,5 kcal/mole.

L'éloignement de ces séquences par rapport au codon ATG nous a conduit à penser que la phase G3 pourrait être un gène dont le codon d'initiation est autre qu'ATG. Il s'agirait d'un codon situé à une distance convenable de l'une ou l'autre des deux séquences SD potentielles. L'utilisation d'un codon d'initiation autre qu'ATG a été rapportée dans le cas du gène *nifD*, codant pour une sous-unité de la nitrogénase de *C.pasteurianum* ; le codon d'initiation est dans ce cas GTG (Wang et al., 1987). Il a été montré également que le gène codant pour le facteur d'initiation de la traduction IF3 possédait un ATT comme codon d'initiation, responsable d'un auto-contrôle de la traduction de ce gène (Butler et al., 1987). En résumé, ou bien le gène commence au niveau de l'ATG, et l'absence d'un site fort de fixation du ribosome implique une faible expression du gène G3, ou bien le codon d'initiation est autre qu'ATG, et dans ce cas pourrait être très différent, comme ATT ou GTT, situés respectivement à 6 nucléotides en aval de SD2 et 7 nucléotides en aval de SD1.

En amont de la phase G4, nous avons identifié, à 6 nucléotides du codon d'initiation, une séquence pouvant jouer le rôle d'une séquence SD, GGAGGT (Fig. 33). L'énergie libre de liaison entre le messenger de ce gène et le rRNA 16S serait de -8,1 kcal/mole. De plus, le codon d'initiation est précédé d'une adénine, déjà remarquée par d'autres auteurs au niveau des gènes de bactéries Gram⁺ (Graves et Rabinowitz, 1986).

En amont de la phase G5, une séquence SD potentielle, ATGAGGTGA, a été identifiée et localisée à 6 nucléotides du codon d'initiation (Fig. 34). L'énergie libre de fixation de cette séquence sur le rRNA 16S serait de -9,3 kcal/mole.

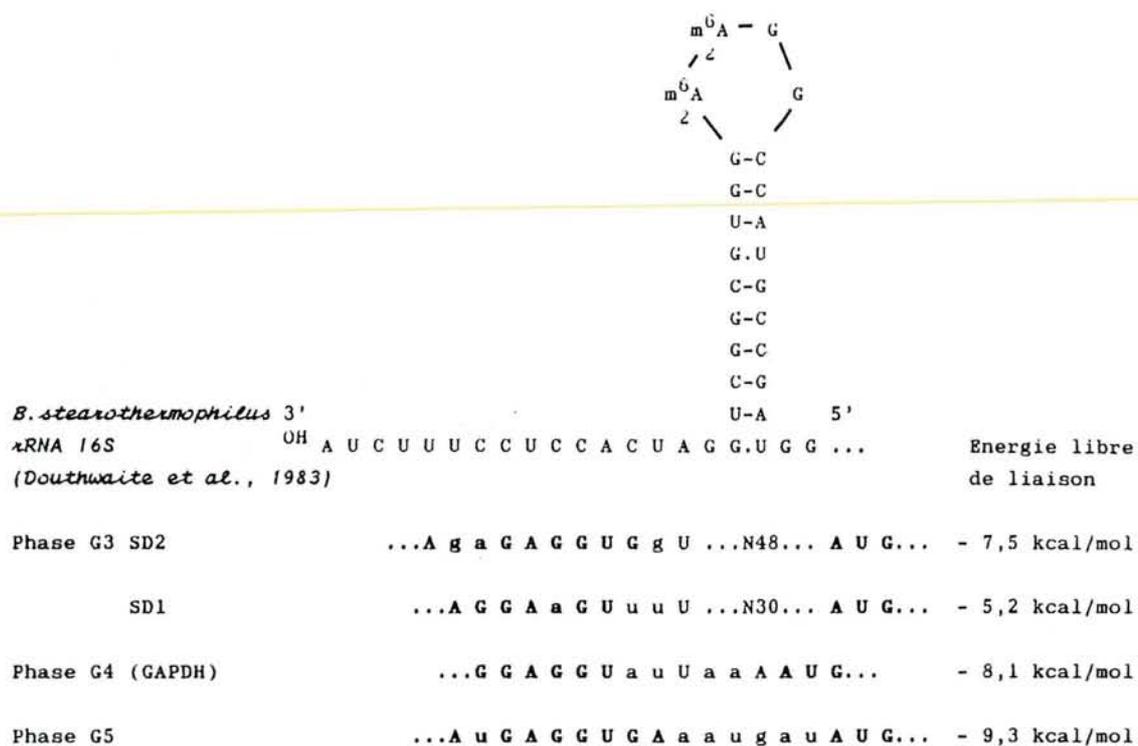


Figure 37 : Les séquences RBS potentielles des gènes de *C. pasteurianum* contenus dans pCGap1

Les séquences RBS sont indiquées en gras, ainsi que les codons d'initiation de la traduction.
 Les nucléotides s'appariant à la séquence du rRNA 16S sont en majuscules.
 (L'énergie libre de liaison avec l'extrémité 3' du rRNA 16S est calculée selon Freier et al., 1986.)

Les séquences Shine-Dalgarno des deux phases ouvertes G4 et G5 présentent un degré élevé de complémentarité avec l'extrémité 3' du rRNA 16S de la bactérie Gram⁺ *B.stearothermophilus* (Fig. 37), confirmant ainsi l'hypothèse de Mc Laughlin et al. (1981) qui avaient suggéré cette forte complémentarité pour l'ensemble des bactéries Gram⁺.

L'étude des signaux d'initiation de la traduction vérifie bien que les phases G4 et G5 sont codantes. Par contre, il est nécessaire de vérifier expérimentalement si la phase G3 l'est également.

III - ETUDE DE LA COMPOSITION EN BASES

Comme nous l'avons déjà mentionné, la composition en bases de l'ensemble du génome de *C.pasteurianum* présente une préférence fortement marquée pour les nucléotides A et T. Le pourcentage en G + C est d'environ 27 % (Cummins et Johnson, 1971). Nous avons obtenu un pourcentage en G + C pour l'ensemble du fragment génomique cloné dans pCGap1 de 28,7 %. L'examen de la composition en bases à l'intérieur de ce fragment montre clairement que les régions codantes se distinguent des régions non codantes par un pourcentage beaucoup plus élevé en G + C.

Les phases ouvertes de lecture G3, G4 et G5 présentent respectivement des pourcentages en G + C de 32, 35 et 32 %. Ceci est un argument supplémentaire en faveur du fait que la phase G3 correspond bien à un gène.

IV - ETUDE DES PHASES OUVERTES DE LECTURE

1. La phase G4

Comme nous l'avons déjà dit, la phase ouverte de lecture G4 correspond au gène de la GAPDH (Fig. 33). Il est à noter qu'elle se situe dans le même cadre de lecture que la phase G3, à 126 nucléotides en aval de son codon de terminaison.

La phase G4 code pour une protéine de 334 amino-acides. Le codon d'initiation est un ATG, le codon de terminaison un TAG.

1-1. Usage des codons

Sur les 61 codons possibles du code génétique, seuls 41 sont utilisés dans le gène de la GAPDH de *C.pasteurianum* (Tab. I). Parmi eux, 27 triplets possèdent un A ou un T en troisième position et codent pour 287 sur 334 amino-acides, soit 85 % de la molécule. *C.pasteurianum* privilégie donc le T et le A en troisième position des codons : le T est présent dans 70% des codons quatre fois dégénérés, tandis que le A l'est dans 87 % des codons six fois dégénérés.

Outre les résidus Met et Trp, quatre amino-acides sont codés presque exclusivement par un seul codon : Cys (TGT), Gln (CAA), Glu (GAA) et Arg (AGA).

Le codon GGC (Gly) est le seul ne renfermant que des nucléotides G et C et n'est utilisé qu'une seule fois.

Le tableau I établit une comparaison de l'usage des codons entre les gènes de GAPDH de *C.pasteurianum* et d'*E.coli*. De nombreux codons sont très utilisés chez *C.pasteurianum*, tandis qu'ils sont peu utilisés dans le gène d'*E.coli* : ATA (Ile), GGA (Gly), AGA (Arg), TTA (Leu) ou TCA (Ser) par exemple. Environ 10 % des codons du gène de la GAPDH de *C.pasteurianum* sont des codons correspondant à des tRNA mineurs chez *E.coli* (Fig. 33).

Nous pouvons donc nous attendre à ce que ce gène nécessite pour son expression des espèces de tRNA mineures chez *E.coli* et soit donc peu exprimé. En effet, Robinson et al. (1984) ont émis l'hypothèse selon laquelle l'usage des codons peut influencer sur l'efficacité de la traduction des gènes dans *E.coli*. Des études plus poussées de Varenne et Lazdunski (1986) suggéraient que l'efficacité de traduction d'un mRNA chez *E.coli* est fortement diminuée si trois codons mineurs se succèdent dans la séquence du messenger. C'est le cas pour le gène de la GAPDH de *C.pasteurianum*, puisque l'on peut observer deux segments de 6 et 5 codons mineurs successifs (Fig. 33), ces deux groupes de codons mineurs étant respectivement situés à 25 et 46 nucléotides en aval du codon d'initiation. Ceci renforçait l'idée que le gène de la GAPDH de *C.pasteurianum* devait être faiblement exprimé dans *E.coli*.

1-2. Structure primaire de la protéine

La structure primaire de la GAPDH de *C.pasteurianum* a été déduite à partir de la séquence nucléotidique de son gène. Elle est représentée sur la figure 33, les amino-acides ayant été numérotés selon Biesecker et al. (1977).

AA	codon	<i>C.past.</i>	<i>E.coli</i>
<i>Met</i>	ATG	6	7
<i>Trp</i>	TGG	2	3
<i>Asn</i>	<u>AAT</u>	<u>11</u>	<u>1</u>
	AAC	7	17
<i>Asp</i>	GAT	15	7
	GAC	4	18
<i>Cys</i>	TGT	1	0
	TGC	2	3
<i>His</i>	CAT	2	1
	CAC	2	5
<i>Gln</i>	CAA	6	0
	CAG	0	5
<i>Glu</i>	GAA	24	13
	GAG	1	2
<i>Lys</i>	AAA	25	26
	AAG	2	1
<i>Phe</i>	TTT	4	1
	TTC	9	10
<i>Tyr</i>	TAT	4	2
	TAC	3	6
<i>Ile</i>	ATT	8	1
	ATC	6	19
<i>Ala</i>	<u>ATA</u>	<u>7</u>	<u>0</u>
	GCT	27	28
	GCC	0	1
	GCA	8	4
	GCG	0	2
<i>Gly</i>	GGT	13	16
	GGC	1	14
	<u>GGA</u>	<u>12</u>	<u>0</u>
	<u>GGG</u>	<u>0</u>	<u>0</u>
<i>Thr</i>	ACT	16	12
	ACC	0	15
	<u>ACA</u>	<u>7</u>	<u>0</u>
	ACG	0	0
<i>Pro</i>	CCT	4	0
	CCC	0	0
	CCA	6	1
	CCG	0	8
<i>Val</i>	GTT	25	21
	GTC	1	1
	GTA	8	8
	GTG	0	4
<i>Arg</i>	CGT	1	8
	CGC	0	4
	<u>CGA</u>	<u>0</u>	<u>0</u>
	<u>CGG</u>	<u>0</u>	<u>0</u>
	<u>AGA</u>	<u>12</u>	<u>0</u>
	<u>AGG</u>	<u>0</u>	<u>0</u>
<i>Leu</i>	CTT	4	0
	CTC	0	0
	<u>CTA</u>	<u>1</u>	<u>0</u>
	CTG	0	19
	<u>TTA</u>	<u>20</u>	<u>1</u>
	TTG	0	0
<i>Ser</i>	TCT	3	6
	TCC	0	9
	<u>TCA</u>	<u>13</u>	<u>0</u>
	TCG	0	0
	AGT	0	0
	AGC	1	0
Total		334	330

Tableau I : Usage des codons dans les gènes de GAPDH de *C.pasteurianum* et d'*E.coli*.
(Les codons soulignés correspondent à des espèces de tRNA mineures chez *E.coli*.)

La séquence protéique présente une forte conservation par rapport à celles des autres GAPDH déjà séquencées. Ceci sera étudié plus en détail dans le chapitre V.

Par rapport à l'enzyme d'*E.coli*, neuf insertions ont été constatées aux positions 140A, 190A, 191A, 263A, 301A, 302A, 304A, 317A et 328A. Par ailleurs, huit délétions ont été localisées aux positions 189, 266, 269, 271-273, 326 et 330. Une grande conservation est observée au niveau des résidus intervenant dans la catalyse ou dans la fixation du NAD⁺, du GAP ou du Pi : Arg 10, Ile 11, Asp 32, Leu 33, Ser 119, Ser 148, Cys 149, Thr 150, His 176, Thr 179, Pro 188, Arg 195, Thr 208, Arg 231, Tyr 311, Asn 313 et Tyr 317.

Il faut souligner également la présence en positions 180 et 185 des résidus Asn et Leu. Ces acides aminés avaient jusqu'à présent été caractérisés uniquement dans les enzymes de bactéries thermophiles. Ceci avait conduit certains auteurs à suggérer que l'existence de liaisons entre l'Asn 180 et la Leu 185 pouvait contribuer à la thermostabilité des GAPDH d'organismes thermophiles (Skarzynski et Wonacott, 1988). Nous reviendrons sur ce point au chapitre V.

2. La phase G5

La phase ouverte G5 est située dans un cadre de lecture différent de celui de la phase G4. Elle est située à 167 nucléotides en aval du codon de terminaison de celle-ci (Fig. 34).

La phase ouverte G5 débute par un codon d'initiation ATG. Comme nous l'avons déjà dit, cette phase est interrompue dans le fragment cloné. Nous ne possédons donc que la région codant pour les 68 premiers acides aminés de l'extrémité N-terminale de la protéine codée par ce gène. L'usage des codons de cette phase est identique à celui du gène de la GAPDH.

Nous avons analysé la structure primaire déduite de la séquence nucléotidique pour essayer d'identifier la protéine qui pourrait être codée par ce gène. Nous avons comparé la séquence obtenue avec celles connues d'autres enzymes glycolytiques. C'est ainsi que nous avons pu identifier la phase G5 comme étant un fragment du gène codant pour l'extrémité N-terminale de la phosphoglycérate kinase (PGK), enzyme intervenant immédiatement après la GAPDH dans la voie glycolytique. Comme la GAPDH, cette enzyme est fortement conservée au cours de l'évolution (Mori et al., 1986).

Conway et Ingram (1988), qui ont séquencé un fragment génomique de *Zymomonas mobilis* renfermant le gène de la GAPDH, ont aussi observé que ce gène était suivi par celui de la PGK dans cet organisme, ces deux gènes étant organisés en tandem dans une structure de type opéron. Dans le cas de *C.pasteurianum* seulement, une séquence promotrice reliée au gène de la PGK a été identifiée entre les deux gènes. Cette séquence pourrait constituer un signal de réinitiation de la transcription pouvant indiquer une régulation possible de la transcription du gène de la PGK, indépendamment de celle du gène de la GAPDH placé en amont de ce signal.

Cette organisation ne se retrouve pas chez *E.coli* où les deux gènes sont distants d'environ 1.300 kb (Bachmann, 1983). *Z.mobilis* étant une bactérie Gram⁺ tout comme *C.pasteurianum*, nous avons pensé que l'organisation des gènes de la GAPDH et de la PGK au sein d'un opéron pouvait être caractéristique des bactéries Gram⁺. Pour confirmer ce point, il était intéressant d'étudier l'organisation des gènes chez une autre bactérie Gram⁺. C'est ce que nous décrivons au chapitre IV où est rapportée l'étude de la séquence d'un fragment génomique de *B.stearothermophilus* renfermant le gène de la GAPDH.

3. La phase G3

La phase ouverte de lecture G3 coderait pour une protéine de 326 amino-acides si l'on considère ATG comme codon d'initiation. Cependant comme nous l'avons vu précédemment d'après l'étude des signaux d'initiation de la traduction, cette protéine pourrait être composée de 340 ou 334 résidus en considérant respectivement ATT ou GTT comme codon d'initiation. Cette phase ouverte se termine par un codon TAG.

La protéine codée par ce gène a fait l'objet de plusieurs tentatives d'identification, consistant en des comparaisons de structures primaires avec la structure des protéines dont les séquences sont répertoriées dans la banque de données UWGC (University of Wisconsin Genetic Computation Group) établie par une équipe de l'Université du Wisconsin et réactualisée en Mai 1987. Nous avons de plus effectué des comparaisons particulièrement approfondies avec les séquences d'autres enzymes glycolytiques. Aucune homologie significative n'a pu être observée avec la protéine codée par la phase G3. Nous n'avons donc pas pu l'identifier et l'avons nommée protéine NI.

L'étude de l'usage des codons, réalisée par Assobhei (1988), montrait clairement que cette phase était composée de 15 % de codons correspondant à des tRNA mineurs chez *E.coli*. Outre les codons rares utilisés également dans les phases G4 et G5, la phase G3 comprend d'autres codons non utilisés dans le gène de la GAPDH d'*E.coli*, tels que AAT (Asn), ATT (Ile), CTT (Leu) ou AGT (Ser).

En d'autres termes, l'usage des codons de la phase G3 est différent de celui des phases G4 et G5. Ceci suggère qu'il pourrait s'agir d'une protéine de régulation peu exprimée chez *C.pasteurianum*.

IV - CONCLUSIONS

Le fragment que nous avons cloné pourrait donc renfermer une partie d'un opéron composé de deux gènes au moins, celui de la GAPDH et celui de la PGK. La séquence en amont correspond vraisemblablement à un troisième gène. Des expériences complémentaires sont nécessaires pour le démontrer.

Etant donné l'usage particulier des codons du gène de la GAPDH de *C.pasteurianum*, il était intéressant d'étudier son expression chez *E.coli*. C'est ce que nous allons décrire dans le chapitre suivant.

Chapitre III :

Etude de l'expression du gène
de la GAPDH de
Clostridium pasteurianum
dans *Escherichia coli*

SOMMAIRE

I - EXPRESSION DU GENE DE LA GAPDH DE <i>C.pasteurianum</i> A PARTIR DES DIFFERENTS PLASMIDES QUE NOUS AVIONS CONSTRUITS.....	page 145
II - SOUS-CLONAGE ET ETUDE DE L'EXPRESSION DU GENE DE LA GAPDH DE <i>C.pasteurianum</i> DANS DES VECTEURS D'EXPRESSION.....	page 146
1. Expression du gène de la GAPDH de <i>C.pasteurianum</i> sous le contrôle du promoteur lacZ	page 147
1-1. Construction du plasmide pCGap5.....	page 147
1-2. Etude de l'expression du gène sous-cloné dans pCGap5.....	page 147
2. Expression du gène de la GAPDH de <i>C.pasteurianum</i> sous le contrôle du promoteur gap d' <i>E.coli</i>	page 149
2-1. Stratégie.....	page 149
2-2. Construction du plasmide pCGap6.....	page 149
2-2-1. Création du site NdeI.....	page 149
2-2-2. Reconstitution du gène muté.....	page 151
2-3. Etude de l'expression du gène sous-cloné dans pCGap6.....	page 154
III - DISCUSSION.....	page 154

Les gènes clonés codant pour la GAPDH d'*E.coli* et pour celle de *B.stearothermophilus* sont fortement exprimés chez *E.coli*, l'enzyme représentant respectivement 80 et 50 % des protéines solubles cytoplasmiques (Branlant et al., 1983). Il était intéressant d'étudier quel serait le taux d'expression pour le gène de *Clostridia* dont l'usage des codons est si particulier. Il avait déjà été montré par ailleurs qu'un certain nombre de gènes de *C.pasteurianum* étaient bien surexprimés chez *E.coli*, notamment le gène de la ferredoxine (Graves et Rabinowitz, 1986) et celui de la protéine mop (Hinton et Freyer, 1986).

Nous avons entrepris d'étudier de façon approfondie l'expression du gène de la GAPDH de *C.pasteurianum* dans *E.coli*.

I - EXPRESSION DU GENE DE LA GAPDH DE *C.pasteurianum* A PARTIR DES DIFFERENTS PLASMIDES QUE NOUS AVIONS CONSTRUITS

Nous avons testé l'expression du gène de la GAPDH de *C.pasteurianum* dans les différentes constructions décrites au chapitre I : pCGap1, le plasmide isolé de la banque (Fig. 25) ; pCGap2, le plasmide obtenu à partir de pCGap1 après élimination du fragment HindIII (Fig. 26) ; pCGap4, le plasmide obtenu par sous-clonage, en aval du promoteur lacZ dans pUC19, du fragment DraI de 1,8 kb de pCGap1 (Fig. 28). Toutes ces constructions ont servi à transformer la souche DF221 dans laquelle toutes ces études ont été réalisées.

Nous avons utilisé comme témoin la souche DF221 transformée par les plasmides pBR322, pF61a (pBR322 recombinant renfermant le gène de la GAPDH d'*E.coli*) et pBst-gap1 (pBR322 recombinant renfermant le gène de la GAPDH de *B.stearothermophilus*). L'activité endogène de la GAPDH d'*E.coli* a été mesurée dans la souche BJ5183 transformée par pBR322.

Les tests d'expression ont été réalisés à partir de cellules transformées, cultivées pendant 24 h dans un volume de 5 ml de milieu M63-succinate-glycérol-ampicilline. L'extrait brut préparé par sonication des cellules a été utilisé pour effectuer deux types d'analyses :

- la mesure de l'activité enzymatique par dosage spectrophotométrique à 340 nm de l'apparition du NADH produit lors de la catalyse enzymatique ;
- la détermination de la quantité relative de GAPDH dans les protéines solubles par étude sur gel de polyacrylamide comme décrit dans "Matériels et Méthodes". L'extrait brut était fractionné sur gel dénaturant. Le gel coloré au bleu de Coomassie était analysé par densitométrie.

Les activités enzymatiques mesurées dans les surnageants de sonication obtenus à partir des cellules transformées par les plasmides pCGap1, pCGap2 et pCGap4 étaient sensiblement équivalentes, environ 3,9 μ moles NADH/ mn/ unité DO₂₈₀. Cette valeur était voisine de celle de l'activité endogène observée dans l'extrait brut préparé à partir des cellules de la souche BJ5183 d'*E.coli* (Tab.II).

L'analyse densitométrique des profils de fractionnement protéique montrait que la GAPDH représentait environ 3 % des protéines solubles, soit une quantité très similaire à celle de la GAPDH d'*E.coli* dans l'extrait préparé à partir de la souche BJ5183 transformée par le plasmide pBR322. En d'autres termes, le gène cloné était faiblement exprimé malgré le grand nombre de copies de plasmides.

Dans chacune des constructions pCGap1 et pCGap2, l'utilisation, pour l'expression du gène de la GAPDH, d'un promoteur appartenant au plasmide était impossible. En effet, le gène était inséré dans le sens opposé de celui du promoteur du gène de résistance à l'ampicilline. D'autre part, l'élimination du promoteur du gène de résistance à la tétracycline dans pCGap2 n'affectait pas l'expression du gène de la GAPDH. C'est pourquoi nous avons pensé que, dans pCGap1 et pCGap2, ce gène s'exprimait vraisemblablement sous le contrôle de son propre promoteur. Le fait d'obtenir la même expression avec pCGap1 et pCGap2 indiquait que le promoteur du gène codant pour la protéine NI (phase ouverte G3), présent dans pCGap1, ne participe pas à l'expression du gène de la GAPDH.

Dans le plasmide pCGap4, le gène est inséré dans un vecteur pUC19, en sens inverse du promoteur lacZ ; il s'exprime donc également à partir de son propre promoteur. Il est par conséquent logique d'observer, là encore, le même taux d'expression.

Il était donc important de tester quelle pouvait être l'expression du gène de la GAPDH lorsqu'il était placé sous le contrôle d'un autre promoteur.

II - SOUS-CLONAGE ET ETUDE DE L'EXPRESSION DU GÈNE DE LA GAPDH DE *C.pasteurianum* DANS DES VECTEURS D'EXPRESSION

Afin de savoir si la faible expression observée pour le gène de la GAPDH de *C.pasteurianum* était due à la faible efficacité de son promoteur ou à son usage particulier de codons, nous avons entrepris de placer ce gène sous le contrôle du promoteur lacZ ou du promoteur gap d'*E.coli*.

1. Expression du gène de la GAPDH de *C.pasteurianum* sous le contrôle du promoteur lacZ

1-1. Construction du plasmide pCGap5

Le gène clostridial, situé dans le fragment DraI de 1,8 kb de pCGap1, avait été sous-cloné dans le site HincII de pUC19 (Fig.28). Le gène, dans pCGap4, était alors orienté en sens inverse du promoteur lacZ. Pour l'orienter dans le même sens, nous avons récupéré le fragment BamHI - HindIII de pCGap4 et l'avons inséré dans un vecteur pUC18 coupé par les mêmes enzymes (Fig.38). Le plasmide pCGap5 ainsi obtenu contenait donc le gène de la GAPDH de *C.pasteurianum* orienté dans un sens tel qu'il puisse être exprimé à partir du promoteur lacZ.

La souche DF221 transformée par le plasmide pCGap5 a pu se développer sur milieu M63-glucose-ampicilline et est donc complétée par ce plasmide.

1-2. Etude de l'expression du gène sous-cloné dans pCGap5

Nous avons procédé comme pour les autres plasmides. Nous avons ainsi observé que l'activité enzymatique de la GAPDH dans le surnageant de sonication était d'environ 19,2 μ moles NADH/mn/ unité DO₂₈₀ (Tab.II). Cette valeur était donc environ 5 fois supérieure à celle observée pour le gène sous contrôle de son propre promoteur. L'analyse densitométrique du gel de fractionnement (Fig.42) confirmait ce résultat, la GAPDH représentant dans ce cas environ 15 % des protéines solubles.

Nous sommes donc parvenus à surexprimer le gène de la GAPDH de *C.pasteurianum* dans *E.coli* en le plaçant sous le contrôle du promoteur lacZ. Il est donc important de noter que cette surexpression a lieu malgré la distance importante entre le promoteur et la partie codante, environ 700 nucléotides (le site DraI qui a servi au sous-clonage dans pUC19 se situe à plus de 650 nucléotides du codon d'initiation), et surtout cette expression assez élevée est obtenue malgré la présence de 10 % de codons correspondant à des tRNA mineurs d'*E.coli*.

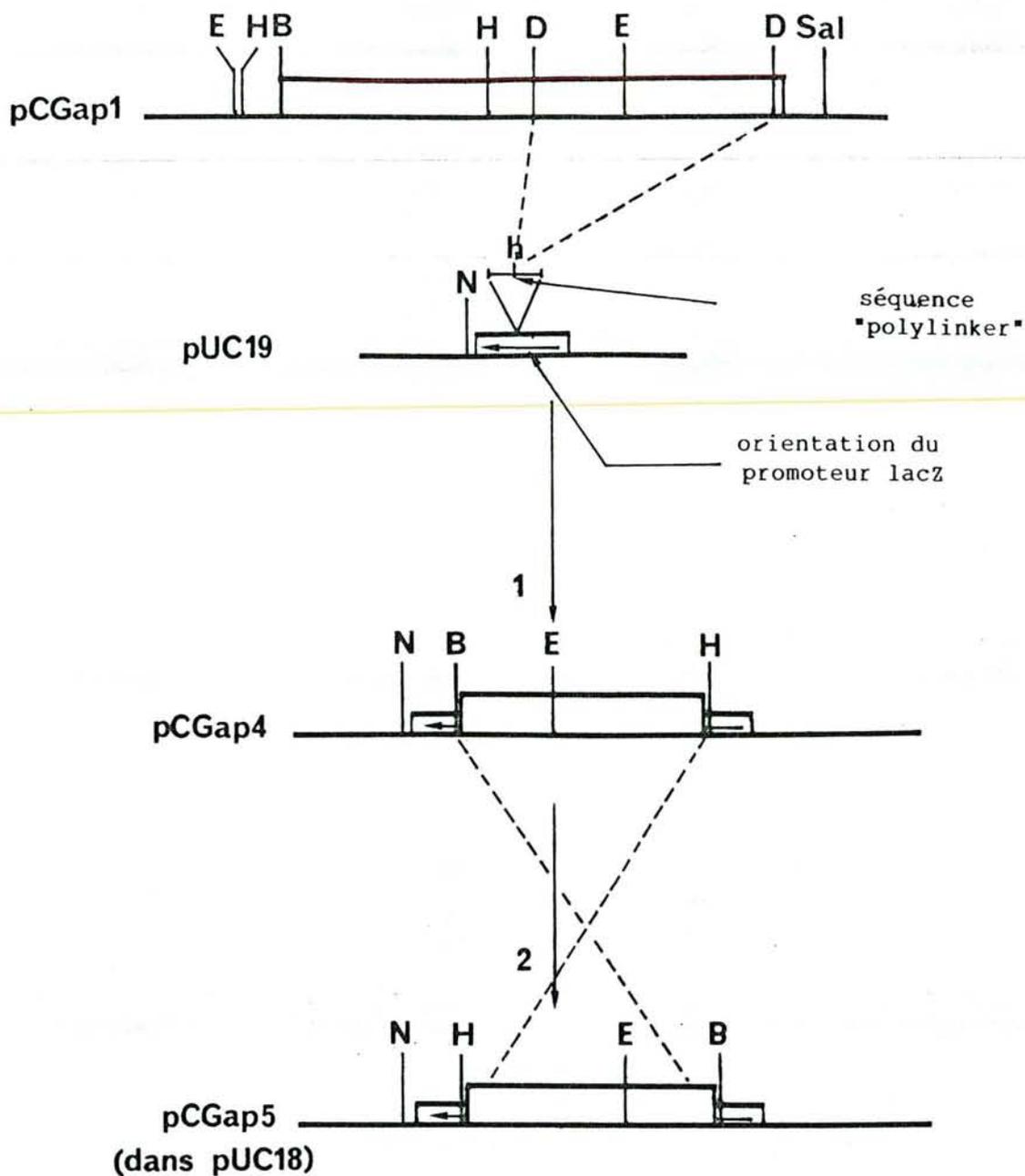


Figure 38 : Construction du plasmide pCGap5

1 : Le plasmide pCGap4 a été obtenu après introduction du fragment DraI de 1,8 kb de pCGap1 dans le site HindII de pUC19. Le gène de la GAPDH est alors orienté en sens inverse du promoteur lacZ.

2 : Afin d'orienter le gène de la GAPDH dans le même sens que le promoteur lacZ, nous avons récupéré le fragment inséré en hydrolysant pCGap4 avec les enzymes BamHI et HindIII. Ce fragment a été réintroduit dans un vecteur pUC18 préparé par les mêmes enzymes. La construction obtenue est appelée pCGap5.

B: BamHI - D: DraI - E: EcoRI - h: HindII - H: HindIII - N: NdeI - Sal: SalI

2. Expression du gène de la GAPDH de *C.pasteurianum* sous le contrôle du promoteur gap d'*E.coli*

2-1. Stratégie

Nous avons déjà évoqué la structure particulière des régions promotrices du gène de la GAPDH d'*E.coli* (Branlant et Branlant, 1985). Cette structure pouvait être l'un des éléments expliquant la surexpression du gène cloné d'*E.coli*.

Au laboratoire, MM. Fournernet et Eghbali ont inséré, dans la séquence "polylinker" du vecteur Bluescript+, le gène de la GAPDH d'*E.coli*, c'est-à-dire la partie codante et les régions non codantes contenant les signaux de régulation de transcription et de traduction de ce gène (Fig.39). Afin de pouvoir utiliser cette construction comme point de départ d'un vecteur d'expression, ils ont créé par mutagenèse dirigée un site NdeI, de séquence CATATG, au niveau du codon d'initiation du gène de la GAPDH d'*E.coli*.

L'intérêt offert par cette construction, appelée Bls-gap, était de pouvoir extraire la région codante du gène de la GAPDH et de lui substituer celle d'un autre gène à surexprimer, possédant lui aussi un site NdeI au niveau de son codon ATG. La construction ainsi préparée permettait d'exprimer ce gène à partir du promoteur gap uniquement, le promoteur lacZ de Bluescript étant orienté en sens inverse.

Nous disposions donc de ce vecteur qui nous permettait de placer la région codante du gène de la GAPDH de *C.pasteurianum* en aval du promoteur gap d'*E.coli*. Pour cela il nous fallait créer, par mutagenèse dirigée, un site NdeI au niveau du codon d'initiation du gène clostridial.

2-2. Construction du plasmide pCGap6

2-2-1. Création du site NdeI

La mutagenèse dirigée a été réalisée sur un fragment du gène de la GAPDH de *C.pasteurianum*, le fragment HindIII - EcoRI de 1 kb contenant l'extrémité 5' de la phase ouverte G4 et sous-cloné dans M13mp9. Nous ne pouvions disposer en effet de l'intégralité du gène dans un vecteur M13, en raison de l'instabilité du fragment EcoRI - SalI déjà décrite précédemment.

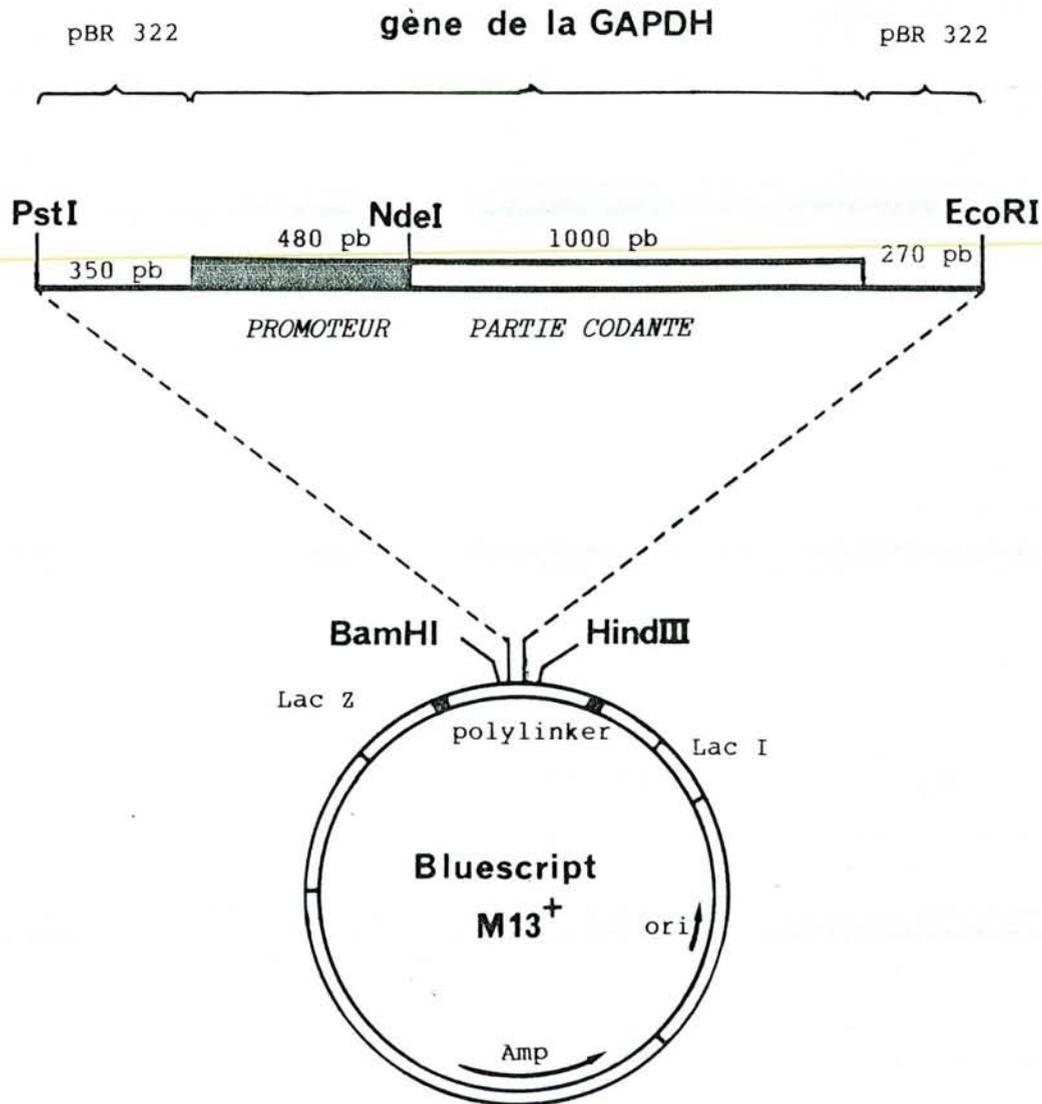


Figure 39 : Carte de restriction du vecteur d'expression Bls-gap

Ce vecteur a été construit au laboratoire à partir d'un plasmide Bluescript dans lequel a été inséré le gène de la GAPDH d'E.coli. Le site NdeI, créé par mutagénèse dirigée, permet de remplacer la partie codante de ce gène par celle du gène à surexprimer.

Nous avons synthétisé un oligonucléotide de 22-mer, porteur des mutations soulignées, ayant la séquence suivante :

3'CCTCCATAAGTATACTGTTTTTC 5'

Cet oligonucléotide permettait donc d'introduire deux mutations au niveau du gène sauvage, la substitution des nucléotides G et A aux deux nucléotides T permettant de créer un site NdeI. Nous avons choisi de synthétiser un oligonucléotide assez long du fait de la richesse en A et T de la séquence environnant les mutations à effectuer.

La mutagénèse a donc consisté en la réalisation d'un hétéroduplex formé d'un brin du phage recombinant M13mp9 contenant le fragment HindIII - EcoRI de *C.pasteurianum* dans le segment "polylinker" hybridé à un brin M13mp19 duquel avait été éliminé le segment "polylinker". La séquence à muter était ainsi située dans un fenêtré en simple brin sur laquelle l'oligonucléotide porteur des mutations pouvait être hybridé (Fig. 22). La mutagénèse a été effectuée dans les conditions décrites dans "Matériels et Méthodes". Les phages mutants ont été sélectionnés dans la souche HB2154 ne possédant pas de tRNA suppresseurs.

La mutagénèse a été réussie dès le premier essai, avec un rendement assez faible toutefois de 25 %. La vérification a été réalisée par séquençage de la région concernée. Parmi les clones négatifs, quelques-uns ne présentaient qu'une seule des deux mutations à effectuer. C'est peut-être ce qui explique le faible rendement obtenu.

2-2-2. Reconstitution du gène muté

La suite des opérations a consisté à introduire le fragment muté dans le plasmide Bls-gap et à reconstituer le gène complet de la GAPDH de *C.pasteurianum* (Fig. 40).

- Dans une première étape, nous avons remplacé le fragment NdeI - EcoRI de Bls-gap, contenant l'intégralité de la région codante du gène de la GAPDH d'*E.coli*, par le fragment NdeI - EcoRI correspondant aux nucléotides codant pour l'extrémité N-terminale de la GAPDH de *C.pasteurianum*.

- Dans une seconde étape, nous avons reconstitué le gène clostridial en insérant le fragment EcoRI - Sall de pCGap3 en 3' du fragment déjà inséré. La construction finale a été contrôlée par coupures par les enzymes de restriction EcoRI, NdeI et Sall (Fig. 41).

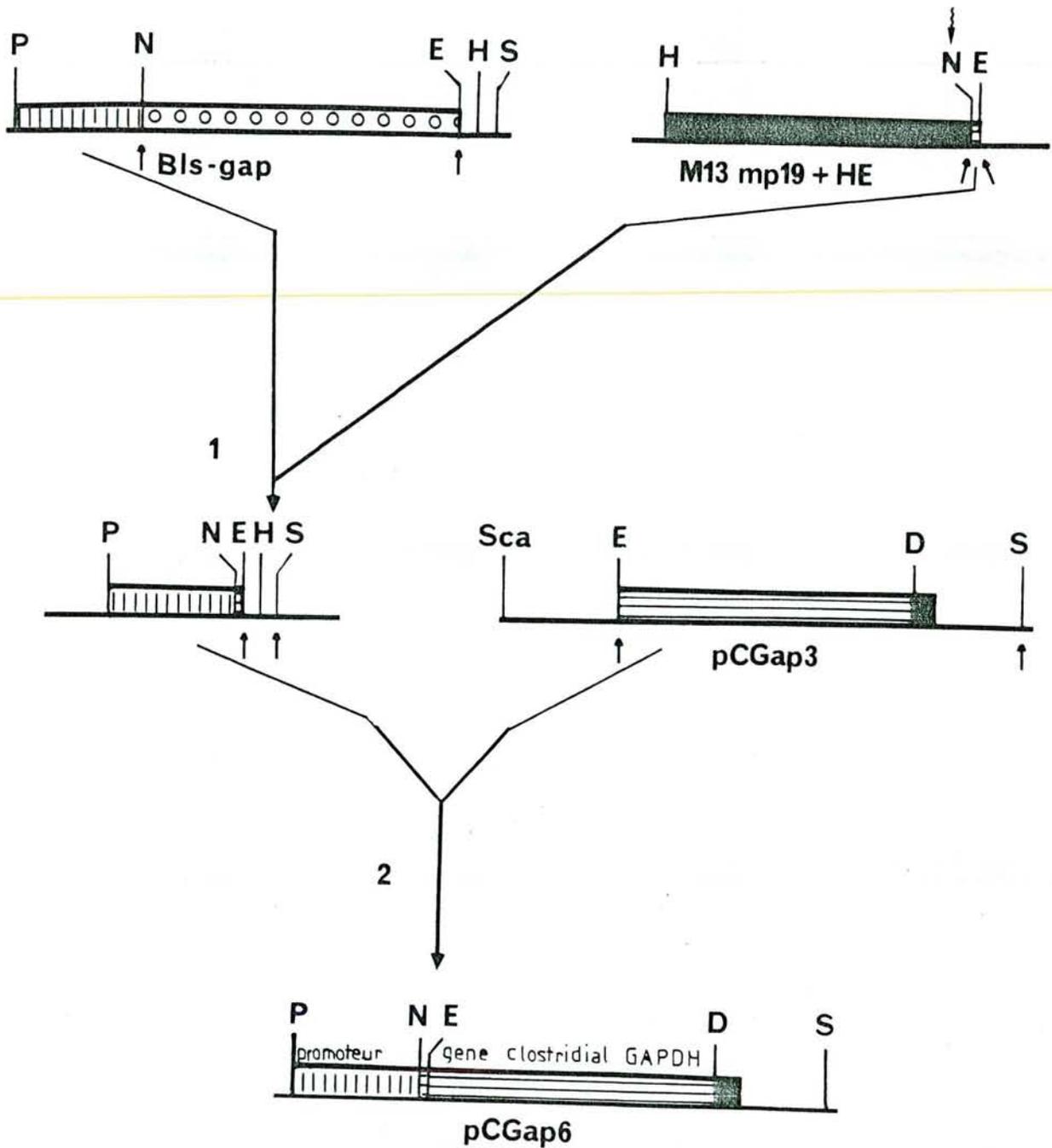


Figure 40 : Construction du plasmide pCGap6

1 : Le fragment HindIII-EcoRI de pCGap1 a été sous-cloné dans M13mp9. Nous avons pu alors créer par mutagenèse dirigée un site NdeI. Le fragment NdeI-EcoRI de Bls-gap a été éliminé et le fragment NdeI-EcoRI présent dans l'insert de M13mp19 lui a été substitué. La construction intermédiaire obtenue contient donc l'extrémité 5' du gène de la GAPDH de *C.pasteurianum* en aval de la région promotrice du gène de la GAPDH d'*E.coli*.

2 : Afin de reconstituer l'intégralité de la région codante du gène clostridial de la GAPDH, le fragment EcoRI-SalI, prélevé à partir de pCGap3, a été inséré dans la construction intermédiaire. La construction obtenue finalement est appelée pCGap6 et contient le gène entier de la GAPDH de *C.pasteurianum* sous le contrôle du promoteur gap d'*E.coli*.

D: DraI - E: EcoRI - H: HindIII - N: NdeI - P: PstI - S: SalI - Sca: ScaI

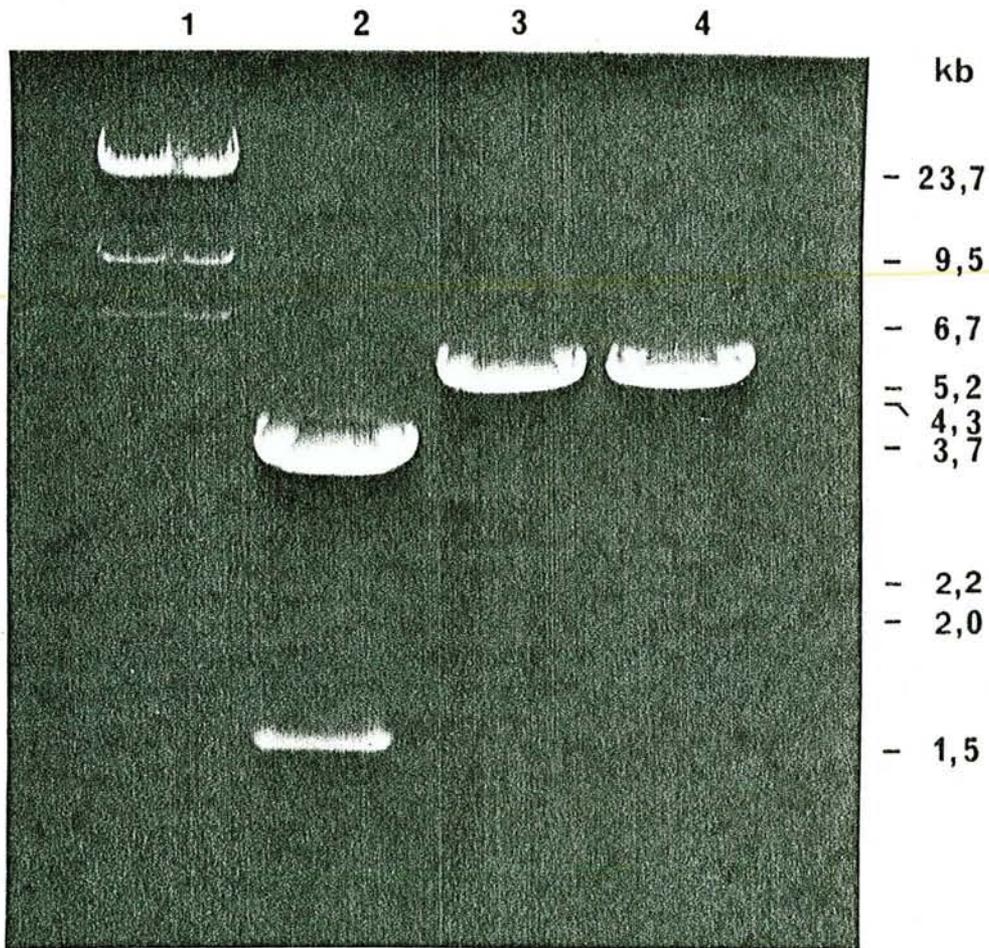


Figure 41 : Fractionnement en gel d'agarose des produits d'hydrolyse de pCGap6

(gel à 0,8 % d'agarose, visualisation du BrEt sous U.V.)

- 1 : DNA du phage lambda / HindIII
- 2 : pCGap6 / EcoRI + Sall
- 3 : pCGap6 / NdeI
- 4 : pCGap6 / EcoRI

Le plasmide pCGap6 obtenu contient donc, dans un vecteur Bluescript, le gène de la GAPDH de *C.pasteurianum* sous le contrôle du promoteur gap du gène de la GAPDH d'*E.coli*. Il complète la souche DF221. De par son orientation, le promoteur lac Z du plasmide Bluescript n'intervient pas dans l'expression du gène de la GAPDH de *C.pasteurianum* qui s'exprime donc à partir du promoteur gap uniquement.

2-3. Etude de l'expression du gène sous-cloné dans pCGap6

Comme l'indique le tableau II, l'activité spécifique de la GAPDH dans le surnageant de sonication est du même ordre de grandeur, 18,3 $\mu\text{moles NADH/ mn/ unité DO}_{280}$, que celle mesurée dans l'extrait protéique des cellules DF221 transformées par le plasmide pCGap5, 19,2 $\mu\text{moles NADH/ mn/ unité DO}_{280}$. De la même façon qu'avec le promoteur lacZ, l'activité enzymatique a été multipliée par environ un facteur de 5 par rapport à celle observée dans la souche DF221 transformée par pCGap1. Le même taux d'expression est donc obtenu avec le gène de la GAPDH de *C.pasteurianum* placé sous le contrôle du promoteur lacZ d'*E.coli* dans pUC18 ou sous le contrôle du promoteur gap d'*E.coli* dans Bls-gap.

Ceci a d'ailleurs été confirmé par l'analyse densitométrique du profil de fractionnement des protéines en gel de polyacrylamide dénaturant (Fig. 42). La GAPDH représente environ 14 % des protéines solubles de l'extrait brut de cellules DF221 transformées par pCGap6, résultat similaire à celui obtenu avec le plasmide pCGap5 (15 %).

III - DISCUSSION

Nous avons constaté que le gène de la GAPDH de *C.pasteurianum* était faiblement exprimé sous le contrôle de son propre promoteur. Nous avons observé cependant que ce gène pouvait être davantage exprimé s'il était placé sous le contrôle d'un autre promoteur comme c'était le cas dans les constructions plasmidiques pCGap5 et pCGap6.

Assobhei (1988) avait mesuré que l'activité spécifique de l'enzyme purifiée de *C.pasteurianum* était de 380 $\mu\text{moles NADH/ mn/ mg d'enzyme}$, équivalente à celle de l'enzyme d'*E.coli* (350 $\mu\text{moles/ nm/ mg d'enzyme}$). Ces valeurs sont bien supérieures à celles de l'activité spécifique de l'enzyme de *B.stearothermophilus* qui est de 100 $\mu\text{moles NADH/ mn/ mg d'enzyme}$. Ceci n'est guère étonnant puisque le test de Ferdinand est réalisé à 37°C, alors que la température optimale de croissance de cette bactérie est de 55°C.

Souches	Plasmides	Activités enzymatiques ($\mu\text{moles NADH/mn/U DO}_{200}$)	Quantités de GAPDH dans l'extrait brut (mg/l de culture)
<i>E. coli</i> BJ5183	pBR322	4,2 \pm 0,8	11
<i>E. coli</i> DF221	pBR322	0,03 \pm 0,01	0,07
<i>E. coli</i> DF221	pCGap1	3,9 \pm 0,9	8,5
<i>E. coli</i> DF221	pCGap2	3,5 \pm 0,8	7,5
<i>E. coli</i> DF221	pCGap4	4,2 \pm 0,5	9
<i>E. coli</i> DF221	pCGap5	19,2 \pm 0,9	42
<i>E. coli</i> DF221	pCGap6	18,3 \pm 1,0	40
<i>E. coli</i> DF221	pBst-gap1 ^a	19 \pm 0,9	100
<i>E. coli</i> DF221	pF61a ^b	91 \pm 5,9	230

^a : pBR322 recombinant contenant le gène de la GAPDH de *B. stearothermophilus*

^b : pBR322 recombinant contenant le gène de la GAPDH d'*E. coli*

Tableau II : Activité spécifique de la GAPDH dans les différentes souches recombinantes DF221

(Les activités enzymatiques indiquées dans ce tableau sont mesurées à 37°C, par le test de Ferdinana, à partir de l'extrait brut obtenu par sonication.)

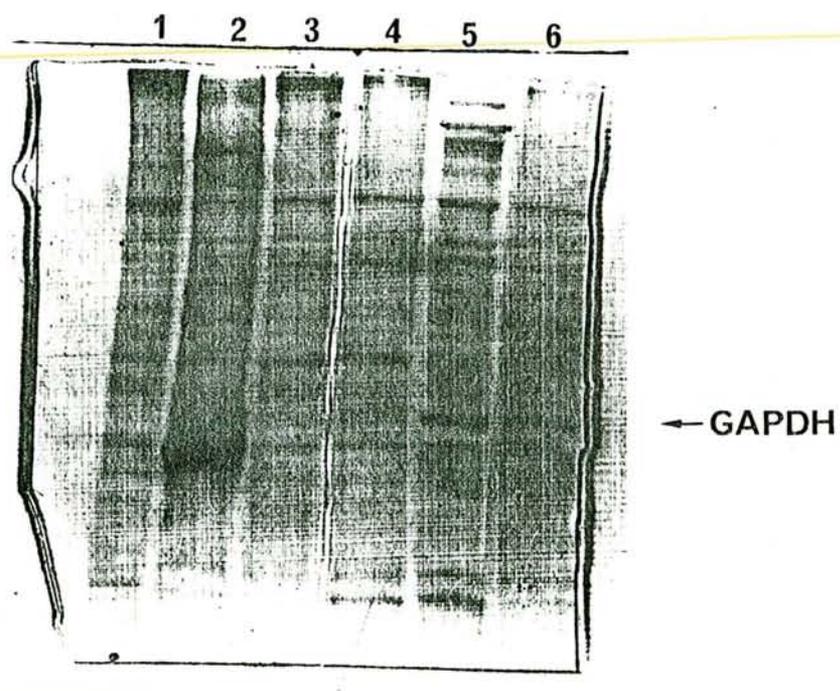


Figure 42 : Fractionnement en gel d'acrylamide des extraits cellulaires des différentes souches recombinantes DF221

(gel à 10 % d'acrylamide dénaturant, révélation au bleu de Coomassie)

1 : DF221 / pBR322
2 : DF221 / pF61a
3 : DF221 / pCGap4

4 : DF221 / pCGap1
5 : DF221 / pCGap6
6 : DF221 / pCGap5

En rapportant l'activité GAPDH mesurée dans l'extrait brut à l'activité spécifique de l'enzyme purifiée, il était possible d'exprimer la quantité d'enzyme présente dans l'extrait. Pour 5 ml de culture, les extraits bruts préparés à partir des cellules DF221 transformées par les plasmides pCGap6, pBst-gap1 et pF61a contenaient respectivement 200, 500 et 1.150 µg de GAPDH de *C.pasteurianum*, de *B.stearothermophilus* et d'*E.coli*. Le tableau II renferme également les quantités d'enzyme obtenues dans l'extrait brut rapportées à des cultures de 1 litre. Ceci montre bien que la surexpression du gène clostridial est bien plus faible que celle des gènes de *B.stearothermophilus* et surtout d'*E.coli*.

En plaçant le gène clostridial sous le contrôle du promoteur gap d'*E.coli*, nous pensons pouvoir obtenir une expression plus forte qu'à partir du promoteur lac Z. Or, le taux d'expression était du même ordre de grandeur.

Deux explications sont possibles :

- le promoteur gap n'est pas plus efficace que le promoteur lac Z, la surexpression du gène cloné d'*E.coli* pourrait donc résulter de l'influence d'éléments autres que les séquences promotrices ;

- le promoteur gap est bien efficace, mais l'usage des codons influencerait sur l'expression du gène clostridial. Il constituerait le facteur limitant, affectant l'efficacité de la traduction et donc la quantité d'enzyme produite dans la cellule. Ceci serait en accord avec l'hypothèse de Varenne et Lazdunski (1986), à savoir que la présence de plusieurs codons rares successifs dans un gène pourrait en limiter l'expression.

Un certain nombre d'expériences restent à réaliser pour déterminer quel est l'élément limitant l'expression du gène de la GAPDH de *C.pasteurianum*, notamment des tests de transcription et de transcription-traduction *in vitro*.

Chapitre IV :

Séquençage et étude d'un fragment de DNA génomique de *Bacillus stearothermophilus* contenant le gène de la GAPDH

Une partie de ce travail a fait l'objet d'un article
actuellement sous presse dans la revue "Gene" et inséré dans ce chapitre.

SOMMAIRE

I - DETERMINATION DE LA SEQUENCE NUCLEOTIDIQUE COMPLETE ET CARACTERISTIQUES GENERALES DU FRAGMENT.....	page 160
1. Séquençage du fragment génomique de 4 kb inséré dans pBst-gap1.....	page 160
2. Caractéristiques générales du fragment séquencé.....	page 162
2-1. La composition en bases.....	page 162
2-2. Les phases ouvertes de lecture.....	page 162
 II - ETUDE DES LOCALISATIONS RESPECTIVES DANS LE GENOME DE <i>B.stearotherophilus</i> DES DIFFERENTS SOUS-FRAGMENTS DU FRAGMENT DE 4 kb INSERE DANS pBst-gap1.....	 page 166
1. Stratégies.....	page 166
1-1. Hybridations de sondes par la technique de Southern.....	page 166
1-2. Etude d'autres plasmides.....	page 168
2. Etude de la région 3'.....	page 168
2-1. Résultats des hybridations.....	page 168
2-2. Etude du plasmide pBst-gap4.....	page 170
3. Etude de la région 5'.....	page 171
 III - ETUDE DETAILLEE DU FRAGMENT DE 3.541 pb.....	 page 171
Publication dans "Gene".....	page 172
Complément à la publication.....	page 183
 IV - ETUDE DE L'EXPRESSION DU GENE DE LA GAPDH A PARTIR DU PLASMIDE pBst-gap4.....	 page 185
 V - LA STRUCTURE EN TANDEM DES GENES DE LA GAPDH ET DE LA PGK.....	 page 186
1. Le promoteur transcriptionnel en amont de l'opéron.....	page 188
2. La région intergénique (GAPDH - PGK).....	page 188
2-1. Les signaux de terminaison de la transcription.....	page 188
2-2. Les signaux d'initiation de la transcription et de la traduction.....	page 188
3. Etude comparative de l'expression des gènes dans <i>E.coli</i>	page 189
 VI - CONCLUSIONS.....	 page 190

M. et Mme Branlant avaient cloné en 1983 le gène de la GAPDH de *B.stearothermophilus* (Branlant et al., 1983). Pour ce faire, une banque de DNA génomique de cet organisme avait été constituée comme décrit précédemment dans le cas de *C.pasteurianum*. Le plasmide pBst-gap1 renfermant un fragment génomique de 4 kb contenant le gène de la GAPDH de *B.stearothermophilus* avait été isolé par complémentation du mutant DF221 GAPDH- d'*E.coli*. D'autres plasmides complétant cette souche avaient été isolés, mais n'avaient pas été analysés de manière approfondie.

Lorsque j'ai entrepris ce travail, le séquençage du fragment de 4 kb inséré dans le plasmide pBst-gap1 était déjà bien avancé. Il restait cependant à lever les ambiguïtés de séquence dans une quinzaine de petits segments. Les difficultés étaient liées à la composition particulièrement riche en G + C du génome de *B.stearothermophilus*. Ce fort pourcentage en G + C entraînait la formation de structures secondaires perturbant la lecture des gels au niveau des séquençages. J'ai alors essayé de résoudre le problème en utilisant un analogue de nucléotide produit par la firme Boehringer Mannheim, le 7-Déaza-dGTP. Cet analogue de nucléotide était décrit comme permettant de réduire les possibilités de formation de structure secondaire, évitant ainsi le tassement des bandes sur le gel de séquençage.

I - DETERMINATION DE LA SEQUENCE NUCLEOTIDIQUE COMPLETE ET CARACTERISTIQUES GENERALES DU FRAGMENT

1. Séquençage du fragment génomique de 4 kb inséré dans pBst-gap1

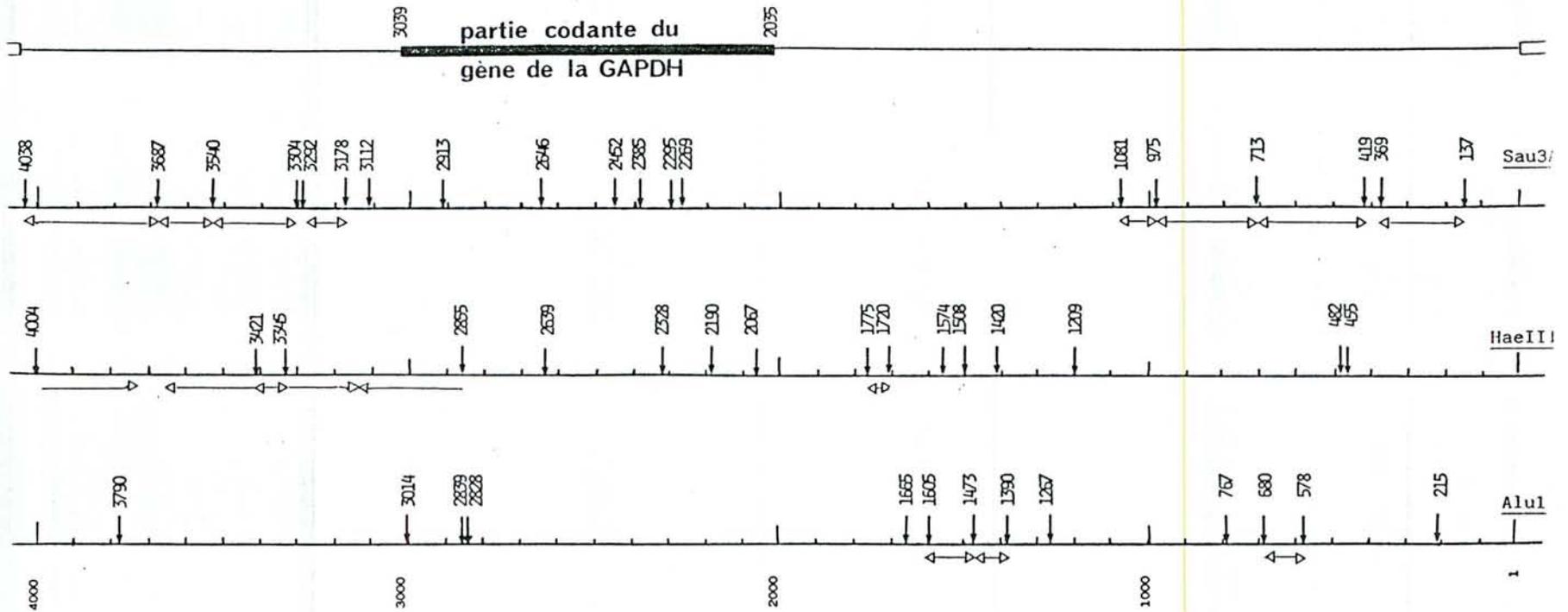
Les fragments qui nécessitaient des vérifications sont indiqués sur la figure 43. Un séquençage répété dans les deux sens de ces différents fragments avec utilisation de 7-déaza dGTP nous a permis de lever les incertitudes de séquence existant dans ces fragments. Nous tenons tout de même à signaler que cet analogue de nucléotide ne représente pas encore la solution parfaite pour séquencer des fragments riches en G + C tel que l'était le fragment génomique de *B.stearothermophilus*.

L'intégralité de la séquence nucléotidique du fragment génomique de 4 kb a été déterminée et est représentée sur la figure 44.

Figure 43 : Localisation des fragments où persistaient des ambiguïtés de séquence au sein du fragment génomique de 4.038 pb cloné dans pBst-gap1

Les flèches noires indiquent les sites de restriction de l'enzyme considérée et les nombres associés leur localisation exacte.

Le sens de lecture est indiqué par la flèche blanche.



2. Caractéristiques générales du fragment séquencé

Le fragment génomique inséré dans pBst-gap1 et contenant le gène de la GAPDH se compose de 4.038 pb. Deux phases ouvertes de lecture ont été identifiées, une grande de 1.005 nucléotides et une petite, située en aval de la grande, de 495 nucléotides (Fig. 44 et 45).

2-1. La composition en bases

Le fragment inséré dans pBst-gap1 présente un pourcentage en G + C d'environ 55 %. A l'inverse de *C.pasteurianum*, les parties codantes de ce fragment génomique de *B.stearothermophilus* ont une composition en bases très similaire à celle des régions non codantes. En effet, les extrémités 5' et 3' du fragment présentent un pourcentage en G + C de 54,5 %, tandis que celui de la grande phase ouverte est de 55,1 %. Nous avons observé, d'après les résultats parus dans la littérature, que le pourcentage en G + C des gènes séquencés variait entre 52 et 58,5 % (Tab. III de la publication). Il était donc nécessaire de vérifier si les phases ouvertes étaient précédées de signaux de transcription et de traduction pour savoir s'il s'agissait bien de régions codantes.

2-2. Les phases ouvertes de lecture

La première, localisée entre les nucléotides 2.035 et 3.042, était la seule dont la taille corresponde à celle attendue pour le gène de la GAPDH. La séquence des 335 amino-acides de la protéine codée par cette phase confirmait bien qu'il s'agissait du gène de la GAPDH.

La seconde phase ouverte s'étendait des nucléotides 3.110 à 3.607 et pouvait coder pour une protéine de 165 amino-acides. Comme nous l'avons mentionné dans le chapitre II, chez *C.pasteurianum* et *Z.mobilis*, les gènes de la GAPDH et de la PGK étaient organisés en tandem. A priori, cela ne semblait pas être le cas chez *B.stearothermophilus*. Mais, en examinant la séquence en amino-acides de la protéine codée par cette seconde phase ouverte de lecture, nous avons observé de fortes homologies avec les séquences N-terminales de la PGK de différents organismes (Mori et al., 1986 ; Conway et Ingram, 1988) : 66 % avec celle de *C.pasteurianum*, 58 % avec celle de *Z.mobilis*, environ 56 % avec celle de *T.brucei* ou de souris. Il était donc étonnant de se trouver face à un gène codant pour une protéine de 165 amino-acides très homologue à la partie N-terminale de la PGK (40 % de la molécule).

Figure 44 : Séquence nucléotidique du fragment de 4.038 pb cloné dans pBst-gap1 et contenant le gène de la GAPDH de *B. stearothermophilus*

	GATCCTGCT	9
ATAATAGAACATGCAACGGAACGAAATAGAGAAATCATGAACTTCCATGCCCTCGTAGTGTAGTGGATAGCACGA		84
GAGATTCGGTTCTCTTAGCGTGGGTTTCAATCCTGCCGAGGGCGTTCCAACGATCTGACAAGTTCGGATTGAGA		159
TTCCCAACCCCTTGATACGACCGCGTTTGCGGTCGTTTTTCATTGCTTGCCGAGCTGGCGGCGGGTTTTCCGC		234
ATCGGAACAAGAGATGGAACGGGGGCTGCGCCTTGTGCAGTCGCTCGACCCGCGGGAGTCGGGCGCCGCGATT		309
GGCGGAATGCCTCTTGCTGCAGCTTGAACGTCTTCTGAGCGCGACGAGCTGGCGGAAACGATCGAGGTGGCTTT		384
CCTCCGCTCAAGTAAAGCGATCATCCAAGAGCTGATCGGGCGGAAGACCGACAGACGCGCTTCTGGCCAAC		459
AGCTTGCTGATTTGCTTCATGGCCGATACGGCATCGCCATTTGCTCCGCACGCGTCGCCAAATACCGCGAGCAG		534
CTGCGCATTCTTCTTCCGGCGAAACGAAAGCAATACACCGAGCTTGATGCGACCGGTGAAAGAAAGGAAGAACG		609
CGATGCACATCCGGTTGTACACGAAGACGAACTGTCCGCTTTGCGACAAAGCGAAAGCGGTTTTGACGGAGCTGC		684
AGGCAGACTATCGGTTTACGATGGAAGAGATCGATATTTATGAAGACGACGCGCTTCTCGAGAAATATCAGTTGA		759
TGATTCGGTCGTCGAGCTTGACGGGGAAGAAATCGGGTATGGGATGATTGAAAAGAAACGGTGAGAAAACGGT		834
TGCGGCGGGCAAAAATA <u>GTTGAAC</u> TGCTATCGTCTCCTGCT <u>TATAAT</u> GAAAATAGCGGGGAGGAATTTTTTAT		909
CGATTGGCGGGACATAATATGTCATACCGGGACAAAAACGTCCCTATGTTCTGCTAAGGGAAAGGATCGATGCA		984
ACCGTTAGTAGAGGCATTAATAAATATCGCCCGACTTGCTTGACGTTATGCAAAAACGATATCAAATTTTGTA		1059
CTCTATTTGCTTATGGCGCCGATCGGCGGGGGCGCTGGCGGCGAGCCTCGGCATGAGCGAGCGGGTTCTCCGG		1134
TCGGAACAGAATTTTGAAGGACAAAACCTTGCTTCCACCGATGTGTGGGCATGCGGTTGACGGAGGAAGGC		1209
CAGGCGCTCCTCATAAGCTCAATGATTTGATGCGGGAAGCGCTCGGACTCAAAGAGCTGGAAGCCGCGCTGCAG		1284
GAACAGCTCGGCATTCGCGGGTGATTGTCGTTGCCGGGATAGCGACCGTCCCTTGGGTAAAAAAGAAATG		1359
GGTAGAGCGTGTGTCGCTGCATGAAGGAGCTGCTTGAAGCCGGGACATCGTCGCGTGGCCGGCGGAACGACGA		1434
TGGCGGCGGTGCGGGAGATGATGACCGCGGATTCGAAGCTTCGTGATGTGTTGTTGTCGCGGCGCGCGGGCC		1509
TTGGCGAAGATGTCGAGAATCAGGCGAACACGATTTGTGCGAAAATGGCGGAAAAGCGATGGGCCGCTATCGGT		1584
TGCTGCAGTGCCTGACCAGCTGAGCGATGAAGCGTACGCGTCGCTTGTGCAAGAGCCGGCGGTGAAAGAAGTGC		1659
TTGAGCTCATCCAATCATGCCGATGGTCGTCCATGGCATCGGTGAAGCGGTGACGATGGCCAAGCGGCGGAAGA		1734
CGCCGCTGTAGAGATGGAGAACATCATCGCCCGCCATGGCCGCGGAGGCTTCGGCTACTATTTCAACGAGTACG		1809
CGGACGTCGTCCATAAAGTGAACAGTGGCATTCAACTTGAACACCTCCCGCATGTCGAGCATGTCATCGCCG		1884

TCGCCGAGGGCGTCGAAGGCGAAAGCATTTCGCGCTACATGAAACGGGCGCCGCACTCGATATTGGTGACGGAC 1959

GAAGGAGCCGCAAAGCGTTTCGTAGGGGAGTAATCTCCCTTGCTGATACAATCCAATCGCAAGGAGGAAAATAAG 2034

ATG GCA GTC AAA GTG GGA ATC AAC GGA TTT GGC CGC ATC GGA CGC AAC GTC TTC CGC 2091
M A V K V G I N G F G R I G R N V F R
1 10

GCG GCA TTG AAA AAC CCG GAC ATT GAA GTG GTG GCG GTG AAC GAT TTA ACC GAT GCG 2148
A A L K N P D I E V V A V N D L T D A
20 30

AAT ACG CTT GCT CAT TTG TTG AAG TAC GAC TCT GTC CAT GGC CGT TTG GAT GCC GAA 2205
N T L A H L L K Y D S V H G R L D A E
40 50

GTG TCG GTG AAC GGC AAC AAC TTG GTC GTC AAC GGC AAA GAA ATC ATC GTC AAG GCG 2262
V S V N G N N L V V N G K E I I V K A
60 70

GAA CGC GAT CCG GAG AAC TTG GCG TGG GGC GAG ATC GGC GTT GAC ATC GTC GTT GAG 2319
E R D P E N L A W G E I G V D I V V E
80 90

TCG ACC GGC CGC TTC ACG AAA CGC GAA GAC GCC GCC AAA CAT TTG GAA GCG GGT GCG 2376
S T G R F T K R E D A A K H L E A G A
100 110

AAA AAA GTG ATC ATT TCC GCC CCG GCG AAA AAC GAA GAC ATT ACG ATC GTC ATG GGC 2433
K K V I I S A P A K N E D I T I V H G
120 122A 130

GTC AAC CAA GAC AAA TAC GAT CCG AAA GCC CAT CAT GTC ATC TCG AAC GCG TCG TGC 2490
V N Q D K Y D P K A H H V I S N A S C
138A 140

ACG ACG AAC TGC TTG GCG CCG TTT GCC AAA GTG CTG CAT GAA CAA TTC GGC ATC GTC 2547
T T N C L A P F A K V L H E Q F G I V
150 160

CGC GGC ATG ATG ACG ACC GTT CAC TCG TAC ACA AAC GAC CAA CGA ATT TTG GAC TTG 2604
R G H H T T V H S Y T N D Q R I L D L
170 180

CCG CAT AAA GAT TTG CGC CGG GCT CGC GCC GCG GCC GAA TCG ATC ATT CCG ACG ACG 2661
P H K D L R R A R A A A E S I I P T T
190 200

ACC GGG GCG GCG AAA GCG GTC GCG CTC GTT TTG CCG GAA TTG AAA GGA AAG TTG AAC 2718
T G A A K A V A L V L P E L K G K L N
210 220

GGC ATG GCG ATG CGC GTG CCG ACG CCG AAC GTA TCC GTT GTC GAC TTG GTG GCG GAA 2775
G H A H R V P T P N V S V V D L V A E
230 240

TTG GAA AAA GAA GTG ACG GTC GAA GAA GTG AAT GCC GCG TTG AAA GCA GCA GCT GAA 2832
L E K E V T V E E V N A A L K A A A E
250 260

GGC GAG CTG AAA GGC ATT TTG GCC TAC AGC GAA GAA CCG CTC GTG TCG CGC GAC TAC 2889
G E L K G I L A Y S E E P L V S R D Y
270 280

AAC GGC AGC ACC GTT TCG TCG ACG ATC GAC GCG TTG TCG ACA ATG GTC ATT GAT GGC 2946
N G S T V S S T I D A L S T H V I D G
290 300

AAA ATG GTG AAA GTC GTT TCG TGG TAT GAC AAC GAA ACG GGC TAT TCG CAC CGC GTC 3003
K H V K V V S W Y D N E T G Y S H R V
310 320

GTC GAC TTA GCT GCC TAC ATC GCC TCG AAA GGG CTG TAA AACCGCTTGGTGCTGAGTTTTCT 3065
V D L A A Y I A S K G L ***

Un examen plus approfondi de l'homologie de séquence montrait qu'elle se limitait aux acides aminés 19 à 145. Au-delà du 145^e résidu, le taux d'homologie chutait considérablement (Fig. 44). Le 145^e acide aminé est codé par les nucléotides 3.542 à 3.544. Or, un site Sau3A est précisément localisé entre les nucléotides 3.542 et 3.545. Rappelons que l'enzyme Sau3A avait été utilisée pour la préparation des fragments et la construction de la banque de *B.stearothermophilus* (Fig. 1 de la publication).

Deux hypothèses étaient envisageables : soit il y avait deux opérons chez *B.stearothermophilus* et dans l'un d'eux le gène de la PGK était dégénéré, soit il y avait eu, lors de la construction de la banque, et plus précisément du plasmide pBst-gap1, ligation de deux fragments génomiques discontinus dans le génome. Le fait que l'homologie s'arrêtait justement au niveau d'un site Sau3A plaidait en faveur de cette hypothèse.

Il était donc nécessaire de tester cette hypothèse. Après avoir observé que le promoteur du gène de la GAPDH était très éloigné de la région codante, il était important de vérifier également si le même événement ne s'était pas produit au niveau de la région 5'-terminale du fragment.

II - ETUDE DES LOCALISATIONS RESPECTIVES DANS LE GENOME DE *B.stearothermophilus* DES DIFFERENTS SOUS-FRAGMENTS DU FRAGMENT DE 4 kb INSERE DANS pBst-gap1

1. Stratégies

1-1. Hybridations de sondes par la technique de Southern

La colinéarité des sous-fragments composant le fragment inséré dans pBst-gap1 a été étudiée par la méthode de Southern. Nous avons choisi de marquer des fragments de restriction appartenant aux régions centrale ou terminales du segment et de les hybrider individuellement à un hydrolysate de DNA génomique fractionné en gel d'agarose puis transféré sur une membrane. L'hydrolyse du DNA génomique a été réalisée par les enzymes EcoRI et BamHI pour lesquelles aucun site n'avait été observé dans l'ensemble du fragment inséré dans pBst-gap1.

En d'autres termes, si tout le DNA de *B.stearothermophilus* inséré dans pBst-gap1 correspondait à des segments colinéaires dans le génome, ces segments devraient être contenus dans un fragment unique après hydrolyse du DNA génomique par les enzymes EcoRI et BamHI.

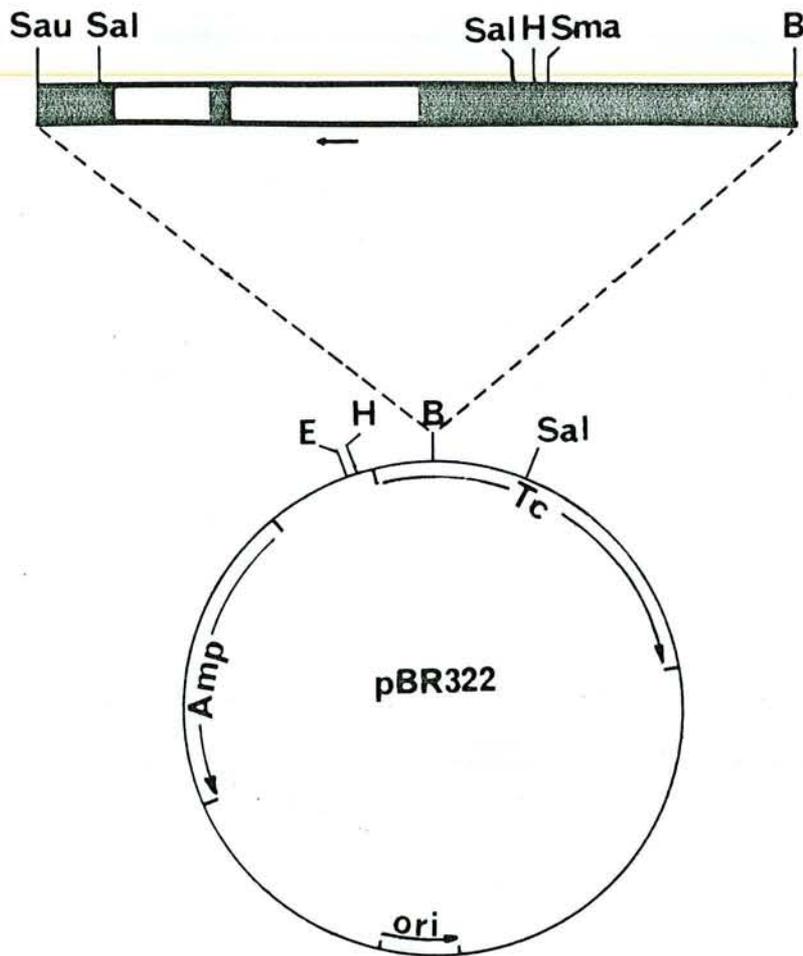


Figure 45 : Carte de restriction du plasmide pBst-gap1.

Les régions codantes sont indiquées en clair.

Nous avons marqué trois sondes de DNA au [^{32}P] par "random priming". Celles-ci ont été préparées à partir de fragments de restriction de pBst-gap1, localisés dans la région codante du gène de la GAPDH (S1), à l'extrémité 3' du fragment génomique inséré de 4.038 pb (S2) et à l'extrémité 5' de ce fragment, c'est-à-dire le segment BamHI - HindIII (S3) (Fig.45).

Le DNA génomique, hydrolysé totalement par les enzymes BamHI et EcoRI, ainsi que les plasmides pBst-gap1 et pBR322 linéarisés par BamHI, ont été fractionnés en gel d'agarose. Les fragments résultants ont été transférés par capillarité sur une membrane de nylon (Hybond-N, Amersham).

Trois membranes ont ainsi été préparées et hybridées avec chacune des trois sondes, comme décrit dans "Matériels et Méthodes".

1-2. Etude d'autres plasmides

Nous disposions d'autres plasmides complétant également la souche DF221 et sélectionnés en même temps que le plasmide pBst-gap1 lors du clonage du gène de la GAPDH de *B.stearothermophilus*. Il était important d'étudier ces plasmides. L'analyse par digestion aux enzymes de restriction Sau3A et HaeIII a montré qu'il y avait deux types de plasmides : des plasmides identiques à pBst-gap1 et des plasmides tous identiques entre eux, ayant une taille d'environ 7,2 kb. Nous avons appelé ces plasmides pBst-gap4.

Après avoir vérifié que tous les plasmides pBst-gap4 libéraient bien les mêmes fragments par hydrolyse par Sau3A et HaeIII, nous avons comparé ces fragments à ceux obtenus avec pBst-gap1.

2. Etude de la région 3'

2-1. Résultats des hybridations

Il est apparu clairement que la sonde S2 et la sonde S1 ne s'hybridaient pas au même fragment de DNA génomique (Fig. 46). En effet, la sonde S2 s'hybridait à un fragment d'environ 2,5 kb, tandis que la sonde S1 s'hybridait, elle, à un fragment d'environ 6 kb. Ceci suggérait donc bien que l'extrémité 3' du fragment de pBst-gap1 n'était pas colinéaire au fragment contenant le gène de la GAPDH dans le génome de *B.stearothermophilus*.

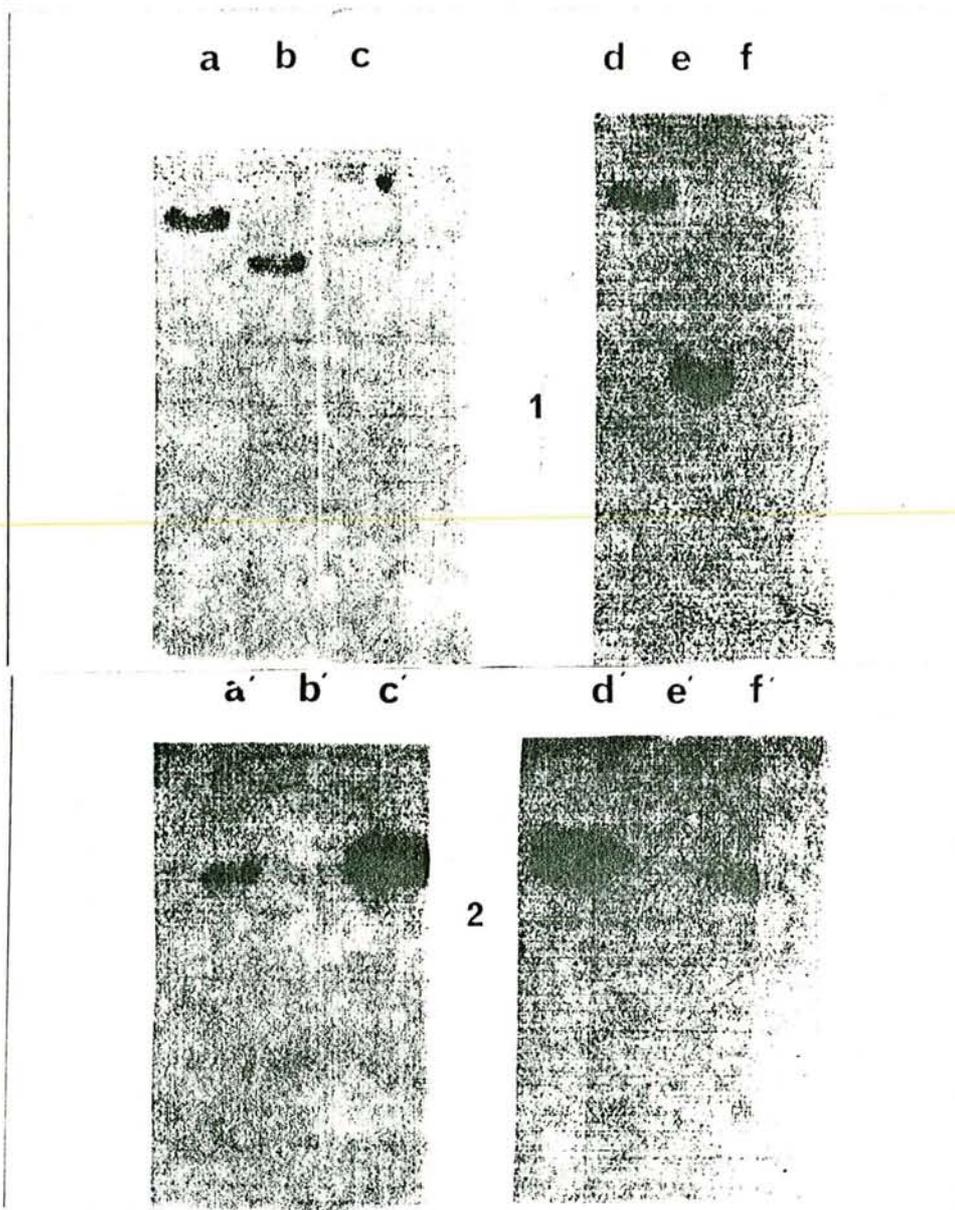


Figure 46 : Hybridations par la méthode de Southern des sous-fragments issus du fragment de 4.038 pb sur le DNA génomique de *B.stearotherophilus*

Les échantillons à tester sont fixés sur des membranes de nylon. Les sondes utilisées correspondent à des fragments purifiés et marqués par la technique de "random priming".

Sonde S1 : sous-fragment issu de la région codante du gène de la GAPDH

Sonde S2 : sous-fragment issu de la région 3' de l'insert de pBst-gap1

Sonde S3 : sous-fragment issu de la région 5' de l'insert de pBst-gap1

1 : Etude de la région 3' du fragment génomique de 4.038 pb.

a, d : pBst-gap1 / BamHI

b, e : DNA génomique de *B.stearotherophilus* / BamHI + EcoRI

c, f : pBR322 / BamHI

Les échantillons a, b et c sont hybridés avec la sonde S1, les échantillons d, e et f avec la sonde S2.

2 : Etude de la région 5' du fragment génomique de 4.038 pb.

a', f' : DNA génomique de *B.stearotherophilus* / EcoRI + HindIII

b', e' : pBR322 / BamHI

c', d' : pBst-gap1 / BamHI

Les échantillons a', b' et c' sont hybridés avec la sonde S1, les échantillons d', e' et f' avec la sonde S3.

2-2. Etude du plasmide pBst-gap4

L'analyse du profil de restriction par Sau3A a montré l'absence de la partie correspondant aux 974 nucléotides en 5' du fragment génomique inséré dans pBst-gap1. Les fragments de la région suivante (975 - 3.541), contenant notamment la région codant pour la GAPDH, étaient tous présents. Les fragments correspondant à la région 3.542 à 4.038 en 3' du fragment génomique inséré dans pBst-gap1 étaient absents. De plus, de nouveaux fragments apparaissaient dans l'hydrolysate par Sau3A du plasmide pBst-gap4, fragments absents dans pBst-gap1. Il nous fallait donc savoir si ces fragments étaient localisés en 5' ou 3' dans le fragment inséré dans le plasmide pBst-gap4.

Nous avons pour cela comparé les profils de restriction des segments HindIII générés à partir des plasmides pBst-gap1 et pBst-gap4 (Fig. 45). Les nouveaux fragments étaient présents dans le segment HindIII préparé à partir du plasmide pBst-gap4. Ceci indiquait que les fragments additionnels de pBst-gap4 étaient situés en 3' et donc que le fragment génomique inséré dans pBst-gap4 présentait une région 3' différente de celle du fragment inséré dans pBst-gap1.

Cette observation confirme les résultats d'hybridation des sondes S1 et S2 sur le DNA génomique. D'ailleurs, la sonde S2, qui correspond à la région 3' du fragment inséré dans pBst-gap1, ne s'hybridait pas au plasmide pBst-gap4.

Comme la probabilité d'avoir systématiquement l'insertion de deux fragments dans un plasmide est faible, nous pensons que la séquence en 3' présente dans le plasmide pBst-gap4 doit correspondre à celle existant réellement dans le génome, deux fragments ayant été mis bout à bout dans pBst-gap1. La colinéarité dans le génome des segments de pBst-gap4 devra tout de même être vérifiée. Il serait d'ailleurs intéressant de déterminer la séquence 3'-terminale du segment contenu dans pBst-gap4, cette séquence devant coder pour la PGK.

En effet, le gène de la GAPDH doit bien être suivi de celui de la PGK dans le génome de *B.stearothermophilus*. Nous retrouvons donc bien, chez *B.stearothermophilus*, l'organisation en tandem des gènes de la GAPDH et de la PGK déjà observée chez les deux autres bactéries Gram⁺ étudiées, *C.pasteurianum* et *Z.mobilis*.

En ce qui concerne le promoteur transcriptionnel situé en amont du gène de la GAPDH, il est, comme nous l'avons déjà signalé, situé à une distance importante du codon d'initiation. Il était donc important de vérifier si, au niveau du génome de *B.stearothermophilus*, il était bien colinéaire à la partie codante. En d'autres termes, il nous fallait vérifier qu'il s'agissait bien du promoteur naturel du gène de la GAPDH de *B.stearothermophilus*.

3. Etude de la région 5'

Comme nous l'avons déjà souligné, la région correspondant aux 974 nucléotides de l'extrémité 5' du fragment inséré dans pBst-gap1 était absente dans le plasmide pBst-gap4. Ce plasmide ne pouvait donc pas apporter d'éléments de réponse pour cette région. La seule approche possible était la technique d'hybridation sur l'hydrolysate de DNA génomique.

Nous avons utilisé les sondes S1 et S3, S1 correspondant à la région codante et S3 au fragment BamHI - HindIII en 5' du gène de la GAPDH (Fig. 45).

L'hybridation a été effectuée sur du DNA génomique coupé par les enzymes BamHI et HindIII. La figure 46 montre que les sondes S1 et S3 s'hybrident exactement au même niveau sur l'hydrolysate, le ou les fragments correspondants ayant une taille approximative de 7,5 kb. Nous ne pouvons pas bien sûr complètement exclure l'hypothèse d'une hybridation sur deux fragments différents ayant exactement la même taille. Néanmoins, la probabilité d'une telle situation est très faible.

Nous pensons donc que les régions 5' et centrale du fragment inséré dans pBst-gap1 sont contiguës et appartiennent au même fragment génomique BamHI - EcoRI.

Pour confirmer ces résultats, la même expérience devra être répétée sur un hydrolysate préparé avec d'autres enzymes de restriction.

III - ETUDE DETAILLEE DU FRAGMENT DE 3.541 pb

Comme nous avons montré que le fragment de 4.038 pb inséré dans pBst-gap1 était composé de deux fragments Sau3A du génome de *B.stearothermophilus*, non colinéaires dans ce génome et longs de 3.541 et 496 nucléotides, nous n'avons pas tenu compte du fragment de 496 nucléotides dans l'article que nous avons rédigé concernant ce travail.

Cet article a été accepté pour publication dans la revue "**Gene**" et est actuellement sous presse.

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*

(Thermophile gene; G + C content; ribosome binding site; transcriptional promoter; terminator; recombinant DNA)

C. Branlant, T. Oster and G. Branlant

Laboratoire d'Enzymologie et de Génie Génétique, UA CNRS 457, Université de Nancy I, Boulevard des Aiguillettes, 54506 Vandœuvre-les-Nancy (France)

Received 14 February 1988

Revised 22 September 1988

Accepted 23 September 1988

Received by publisher 10 October 1988

SUMMARY

The complete nucleotide sequence of a 3541-base pairs (bp) fragment from *Bacillus stearothermophilus* able to complement an *Escherichia coli* glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) mutant (*gapD*⁻) has been determined. The *B. stearothermophilus gap* gene consists of a 1005-bp open reading frame commencing with an ATG start codon and ending with a TAA stop codon. Upstream from the start codon is a strong Shine-Dalgarno sequence typical of Gram-positive bacteria. Only one putative RNA polymerase recognition signal (-35 and -10 regions) is found 1153 bp upstream from the ATG start codon. In vivo utilization of this signal is in agreement with the study of gene expression from different subclones of the original fragment. 57 bp downstream from the TAA stop codon is a structure resembling Rho-independent transcription termination signals. Although *B. stearothermophilus* GAPDH-coding gene is highly expressed in *E. coli*, it contains several rare codons for *E. coli*. The predicted amino acid sequence of the GAPDH enzyme presents several differences with the amino acid sequence previously determined from the protein and is in better agreement with published crystallographic data.

Correspondence to: Dr C. Branlant, Laboratoire d'Enzymologie et de Génie Génétique, Université de Nancy I, Bld. des Aiguillettes, 54506 Vandœuvre-les-Nancy (France) Tel. 83.91.20.97.

gapD, gene coding for *E. coli* GAPDH; NAD, nicotinamide adenine dinucleotide; nt, nucleotide(s); ORF, open reading frame; RBS, ribosome-binding sequence; Tc, tetracycline.

Abbreviations: aa, amino acid(s); Ap, ampicilline; bp, base pair(s); GAPDH, glyceraldehyde-3-phosphate-dehydrogenase;

INTRODUCTION

Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a key enzyme involved in glycolysis and gluconeogenesis. It reversibly catalyzes the oxydative phosphorylation of D-glyceraldehyde-3-phosphate to 1,3-diphosphoglycerate using NAD as a coenzyme.

A strong conservation is observed, at both the primary and the tertiary structure levels, among all GAPDH described up to now (Branlant and Branlant, 1985; Walker et al., 1980).

We previously cloned the *Escherichia coli* and *B. stearothermophilus* GAPDH-coding genes and found that they were highly expressed in *E. coli* (Branlant et al., 1983). Primary structure analysis of the *E. coli* gene revealed the presence of a peculiar structure in the promoter region and a codon usage similar to that for highly expressed protein of *E. coli* (Branlant and Branlant, 1985). It was therefore interesting to compare these peculiar features of the *E. coli* gene to those of the *B. stearothermophilus* gene.

Furthermore, since the three-dimensional structures of *B. stearothermophilus* GAPDH holo- and apoenzymes have been studied at high resolution using x-ray crystallography (Leslie and Wonacott, 1983 and 1984; Skarzynski et al., 1987), knowledge of the primary structure of its gene provided an opportunity to undertake a thorough study of the GAPDH structure-function relationship, using site-directed mutagenesis. Actually, a series of mutants affecting amino acids directly implicated in the catalysis or involved in substrate or coenzyme binding has by now been constructed in the laboratory.

In this paper, we determined the sequence determination of the *B. stearothermophilus* GAPDH-coding gene and its flanking regions. The deduced amino acid sequence is compared to the one previously observed by protein sequencing (Biesecker et al., 1977). The differences observed are discussed in relation with the crystallographic model determined by Skarzynski et al. (1987).

We also describe experiments showing the need for a long stretch of upstream chromosomal DNA sequence for expression of the *B. stearothermophilus* GAPDH-coding gene in *E. coli*. The complete nucleotide sequence of the 3541-bp fragment required for expression in *E. coli* is given. The results obtained

are compared to other known nucleotide sequences of *B. stearothermophilus* genes.

EXPERIMENTAL AND DISCUSSION

(a) Sequence determination and characteristic features of the GAPDH coding region

The strategy used to sequence the 3541-bp fragment of *B. stearothermophilus* chromosomal DNA, inserted in the plasmid pBR322/BstGAPDH (Branlant et al., 1983), is summarized in Fig. 1. For reasons of clarity this plasmid will be named plasmid pBst-gap1 in the present paper. The sequence obtained is shown in Fig. 2; 95% has been determined on both strands. The *B. stearothermophilus* GAPDH-coding gene consists of an ORF of 1005 bp beginning with an ATG start codon and ending with a TAA stop codon (Fig. 2).

(1) Protein sequence

As shown in Table I, 20 aa residues differ from the published amino acid sequence, although the same strain, NCA 1503, was used in both studies (Biesecker et al., 1977). Some of these discrepancies, e.g., Asp → Asn, could result from the difficulties in differentiating these residues by the usual chemical sequencing method. One argument in favor of the validity of the amino acid sequence deduced from the nucleotide sequence is that it fits better with crystallographic data than does the amino acid sequence directly determined for the protein. Recently, Skarzynski et al. (1987) have published a refined version of the crystallographic model taking into account the results we obtained from DNA sequencing. As shown by these authors, 3 of the 20 changes (replacement of Asn-125 by Asp, Asn-181 by Asp, and Gly-195 by Arg) led to important structural and mechanistic consequences. First, two new ion pairs could be formed: Arg-102-Asp-125 and Asp-181-Arg-195, that could stabilize the tertiary structure of the *B. stearothermophilus* enzyme. Secondly, Arg 195, by its positive charged group, could stabilize the binding of phosphate groups by the intermediary of a water molecule and so could contribute to the formation of the phosphate ion binding site in the catalytic domain.

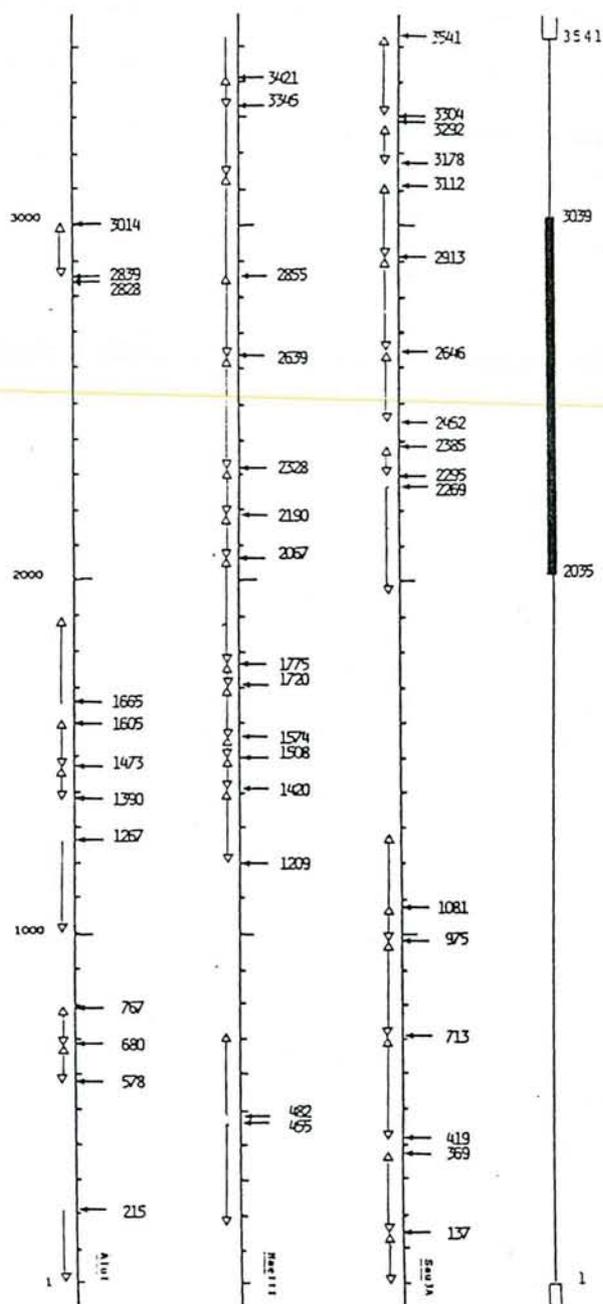


Fig. 1. Restriction sites and sequencing strategy used to establish the nucleotide sequence of plasmid pBst-gapI insert.

Nucleotides are numbered starting from the end of the insert which is proximal to the N-terminus of the gene. The double lines at the ends represent pBR322 sequences, the heavy black line represents the coding region of GAPDH-coding gene and the thin line the flanking regions. Restriction fragments derived from *Sau3A*, *HaeIII* and *AluI* digestions of the plasmid pBst-gapI were purified by electrophoresis on 8% polyacrylamide gels, those corresponding to the insert were subcloned into the phage M13mp9 and sequenced by the dideoxy chain termination method (Sanger et al., 1977). Due to difficulties in purifying all

(2) Codon usage

Table II contains a comparison of the codon utilization for GAPDH of *E. coli* and *B. stearothermophilus*. Strong differences are observed, corresponding to a marked preference for G or C residues at the third position of the codons in *B. stearothermophilus* (71.1% against 51.8% for *E. coli*) (Table III). As can be seen on Tables II and III, such a preference was also observed in other thermophilic genes for which the DNA sequence has been determined (for review, Barstow et al., 1986a), and has been interpreted in relation with the higher optimal growth temperature of thermophilic prokaryotes.

TABLE I

Differences between the *B. stearothermophilus* GAPDH sequences deduced from protein analysis (Biesecker et al., 1977) and from nucleotide sequence of the cloned gene (Branlant et al., 1983)

Amino acid number	From nucleotide sequence	Direct amino acid sequence
36	Asp	Asn
38	Asn	Asp
39	Thr	Gly
58	Ser	Val
61	Gly	Asp
62	Asn	Gly
63	Asn	Asp
64	Leu	Val
65	Val	Ser
78	Asp	Asn
123	Asn	Val
125	Asp	Asn
128	Ile	Val
163	Glu	Gln
164	Gln	Glu
181	Asp	Asn
195	Arg	Gly
282	Asp	Asn
329	Ala	Asn
330	Ser	Ala

subfragments starting from the plasmid pBst-gapI digests, we used the recombinant phage M13mp9 having inserted the *AluI* fragment (nt 1665-2828) as an intermediate. This phage was digested by *Sau3A* and by *HaeIII* and the resulting fragments corresponding to the insert were subcloned in phage M13mp9 and sequenced. The direction and extent of sequence determination are shown by horizontal arrows. In most cases, the nucleotide sequence has been determined on both strands.

Several of the frequently used codons in the *B. stearothermophilus* GAPDH-coding gene correspond to codons rarely found in highly expressed *E. coli* genes, e.g., UUG (Leu, GUC (Val), ACG

(Thr) and UCG (Ser), the latter being considered to be a rare codon in general in *E. coli* (Sharp and Wen-Hsiung Li, 1986). Interestingly, in spite of this codon usage, a high expression level of the cloned

A	GATCCTGCT	9
ATAATAGAACATGCAACGGAAACGAAATAGAGAAATCATGAACTTCCATGCCCTCGTAGTGTAGTGGATAGCACGA		84
GAGATTCGGTTCTCTTAGCGTGGGTTCTGAATCCTGCCGAGGGCGTTCCAACGATCTGACAAGTTCGGATTGAGA		159
TTCCCAACCCCTTGATACGACCGCGTTTTCGGTCTTTTTTTCATTGCTTGGCGAGCTGGCGGCGGGTTTTCCGC		234
ATCGGAACAAGAGATGGAACGGGGGCTGCCCTTGTGCAGTCGCTCGACCCCGGGGAGTCGGCGCCCGGATT		309
GGCGGAATGCCTCTTGCTGCAGCTTGAACGCTTCTCCTGAGCGGACGAGCTGGCGGAAACGATCGAGGTGGCTTT		384
CCTCCGCTCAAGTGAAGCGATCATCCAAGAGCTGATCGCGGCGAAGACCGACAGACCGCTTTCTGGCCAAC		459
AGCTTGCTGATTGTCTTCATGGCCGATACGGCATCGCCATTTTCGCTCCGACGGGTCCGCAAATACGGCGAGCAG		534
CTGCGCATTCTTCTTTCGGCGAAACGAAAGCAATACACCGGAGCTTGTGCGACCGGTGAAAGAAAGGAAGAACG		609
CGATGCACATCCGGTTGTACACGAAGACGAACTGTCCGCTTTGCGACAAAGCGAAAGCGGTTTTGACGGAGCTGC		684
AGGCAGACTATCGGTTTACGATGGAAGAGATCGATATTTATGAAGACGACGCGCTTCTCGAGAAATATCAGTTGA		759
TGATTCGGTCTGTCGAGCTTACCGGGAAGAAATCGGGTATGGGATGATTGAAAAAGAAACGGTGAGAAAAACGGT		834
TGCGGGGGGCAAAAAATGTTGAACGCTATCGTCCTCCTGCTATAATGAAAAATAGCGGGGAGGAATTTTTTAT		909
CGATTGGCGGACATAATATGTACATACCGGACAAAAACGTCCTATGTTCTGCTAAGGGAAGGATCGATGCA		984
ACCGTAGTAGAGGCATTAATAAATATCGCCGACTTGGCTTACGCTTATGCAAAAAACGATATCAAATTTTGT		1059
CTCTAATTCGCTTATGGCGCCGATCGGCGGCGGGCGTGGCGGCGAGCCTCGGCATGAGCGAGCGGGTTCTCCGG		1134
TGGAAACAGAATTTTTGAAAGGACAAAACCTGCTTCCACCGATGTGTCGGGCATGCGGTTGACGGAGGAAGGC		1209
CAGGCGCTCCTTCATACGCTCAATGATTTGATGCGGGAAGCGCTCGGACTCAAAGAGCTGGAAGCCGCGCTGCAG		1284
GAAACAGCTCGGCATTCGCGGGTGTGTCGTTGCGGGGATAGCGACCGTCCCTTGGGTAAAAAAGAAATG		1359
GGTAGAGCGTGTGTCGCTGCATGAAGGAGCTGCTTGAGCCCGGGACATCGTCGCGTGGCCGGCGGAACGACGA		1434
TGGCGGCGGTGCGGAGATGATGACCGCGATTGAAAGCTTCGTGATGTGTTGTTTCGTGCCGCGCGCGGGGCC		1509
TTGGCGAAGATGTGCGAATCAGGCGAACACGATTTGTGCGAAAATGGCGGAAAAACGATGGGCCGCTATCGGT		1584
TGCTGCACGTGCTGACCAGCTGAGCGATGAAGCGTACCGCTCGCTTGTGCGAAGAGCCGGCGGTGAAAGAAGTGC		1659
TTGAGCTCATCCAATCATGCCGATGGTCCATGGCATCGGTGAAGCGGTGACGATGGCCAAGCGCGGAAGA		1734
CGCCGCTGTAGAGATGGAGAACATCATGCCCGCATGGCCGCGAGGCTTCGGCTACTATTTCAACGAGTACG		1809
GCGACGTCGTCCATAAAGTAAAAACAGTCGGCATTCAACTTGAACACCTCCCGCATGTCGAGCATGTCATCGCGG		1884
TGCGCGGAGGGCGTGAAGGCGAAAGCATTGCGCGTACATGAAACGGGCGCGCACTCGATATTGGTGACGGAC		1959

Fig. 2a.

B	GAAGGAGCCGCAAAAAGCGTTCGTAGGGGAGTAATCTCCCTTGCTGATACAATCCAATCGCAAGGAGGAAAAATAAG	2034
	ATG GCA GTC AAA GTG GGA ATC AAC GGA TTT GGC CGC ATC GGA CGC AAC GTC TTC CGC	2091
	H A V K V G I N G F G R I G R N V F R	
	1 10	
	GCG GCA TTG AAA AAC CCG GAC ATT GAA GTG GTG GCG GTG AAC GAT TTA ACC GAT GCG	2148
	A A L K N P D I E V V A V N D L T D A	
	20 30	
	AAT ACG CTT GCT CAT TTG TTG AAG TAC GAC TCT GTC CAT GGC CGT TTG GAT GCC GAA	2205
	H T L A H L L K Y D S V H G R L D A E	
	40 50	
	GTG TCG GTG AAC GGC AAC AAC TTG GTC GTC AAC GGC AAA GAA ATC ATC GTC AAG GCG	2262
	V S V N G N N L V V N G K E I I V K A	
	60 70	
	GAA CGC GAT CCG GAG AAC TTG GCG TGG GGC GAG ATC GGC GTT GAC ATC GTC GTT GAG	2319
	E R D P E N L A W G E I G V D I V V E	
	80 90	
	TCG ACC GGC CGC TTC ACG AAA CGC GAA GAC GCC GCC AAA CAT TTG GAA GCG GGT GCG	2376
	S T G R F T K R E D A A K H L E A G A	
	100 110	
	AAA AAA GTG ATC ATT TCC GCC CCG GCG AAA AAC GAA GAC ATT ACG ATC GTC ATG GGC	2433
	K K V I I S A P A K H E D I T I V H G	
	120 122A 130	
	GTC AAC CAA GAC AAA TAC GAT CCG AAA GCC CAT CAT GTC ATC TCG AAC GCG TCG TGC	2490
	V N Q D K Y D P K A H H V I S N A S C	
	138A 140	
	ACG ACG AAC TGC TTG GCG CCG TTT GCC AAA GTG CTG CAT GAA CAA TTC GGC ATC GTC	2547
	T T N C L A P F A K V L H E Q F G I V	
	150 160	
	CGC GGC ATG ATG ACG ACC GTT CAC TCG TAC ACA AAC GAC CAA CGA ATT TTG GAC TTG	2604
	R G H M T T V H S Y T N D Q R I L D L	
	170 180	
	CCG CAT AAA GAT TTG CGC CCG GCT CGC GCC GCG GAA TCG ATC ATT CCG ACG ACG	2661
	P H K D L R R A R A A E S I I P T T	
	190 200	
	ACC GGG GCG GCG AAA GCG GTC GCG CTC GTT TTG CCG GAA TTG AAA GGA AAG TTG AAC	2718
	T G A A K A V A L V L P E L K G K L N	
	210 220	
	GGC ATG GCG ATG CGC GTG CCG ACG CCG AAC GTA TCC GTT GTC GAC TTG GTG GCG GAA	2775
	G H A H R V P T P N V S V V D L V A E	
	230 240	
	TTG GAA AAA GAA GTG ACG GTC GAA GAA GTG AAT GCC GCG TTG AAA GCA GCA GCT GAA	2832
	L E K E V T V E E V N A A L K A A A E	
	250 260	
	GGC GAG CTG AAA GGC ATT TTG GCC TAC AGC GAA GAA CCG CTC GTG TCG CGC GAC TAC	2889
	G E L K G I L A Y S E E P L V S R D Y	
	270 280	
	AAC GGC AGC ACC GTT TCG TCG ACG ATC GAC GCG TTG TCG ACA ATG GTC ATT GAT GGC	2946
	N G S T V S S T I D A L S T H V I D G	
	290 300	
	AAA ATG GTG AAA GTC GTT TCG TGG TAT GAC AAC GAA ACG GGC TAT TCG CAC CGC GTC	3003
	K H V K V V S W Y D N E T G Y S H R V	
	310 320	
	GTC GAC TTA GCT GCC TAC ATC GCC TCG AAA GGC CTG TAA AACCAGCTTGGTGTGAGTTTC	3065
	V D L A A Y I A S K G L ***	
	330 333	

Fig. 2b.

C	CGCCTGTTCAAGTCTATATAGGAAAATGGAGGG	<u>GAGCGGGGAAATG ATC CCG ACT CCT TTT</u>	CCT TGT	3133
			<u>H I P T P F P C</u>	
CCA AAC TGT AGC CAA AAG GGG GCA CGA ACG ATG AAC AAG AAG ACG ATC CGC GAC GTT				3190
F N C S Q K G A R T H H K K T I R D V				
GAT GTG AGG GGA AAG CGC GTC TTT TGC CGC GTC GAT TTC AAC GTT CCG ATG GAG CAA				3247
D V R G K R V F C R V D F N V P H E Q				
GGC GCT ATC ACC GAT GAC ACG CGC ATT CUC GCC GCA UTC CCG ACG ATC CGC TAT TTG				3304
G A I T D D T S I R A A L P T I R Y L				
ATC GAG CAC GGG GCG AAA GTC ATT TTG GCG AGC CAC CTC GGC CGC AAA GGA AAA				3361
I E H G A K V I L A S H L G R P K G K				
GTG GTC GAA GAA TTG CGT TTG GAT GCC GTT GCG AAG CGG CTC GGC GAG CTG CTT GAA				3418
V V E E L R L D A V A K R L G E L L E				
CGG CCG GTT GCC AAA ACG AAT GAA GCG GTC GGC GAT GAG GTG AAA GCG GCG GTC GAC				3475
R P V A K T N S A V G D E V K A A V D				
CGT TTG AAC GAA GGC GAT GTG CTC TTG CTT GAG AAC GTC CGT TTT TAC CCT GGC GAA				3532
R L H E G D V L L L E N V R F Y P G E				
GAG AAA AAT				3541
E K N				

Fig. 2. Complete nucleotide sequence of plasmid pBst-gap1 insert represented in Fig. 1 and the predicted protein sequence of *B. stearothermophilus* GAPDH.

Initiation codon (ATG), termination codon (TAA) and Shine-Dalgarno complementary sequence are underlined. Putative transcription promoter and terminator are boxed. The numbering of the amino acid sequence was according to Biesecker et al. (1977).

B. stearothermophilus GAPDH-coding gene was observed in *E. coli* (40% of the soluble proteins). This observation is in agreement with recent data suggesting that reduction of translation rate by means of 'rare' codon utilization is not highly effective, except when three or more 'rare' codons are present side by side (Robinson et al., 1984; Bonekamp et al., 1985; Varenne and Lazdunski, 1986).

(b) Features of DNA regions upstream and downstream from the open reading frame

The sequence analysis presented in this paper confirms that in plasmid pBst-gap1, the *B. stearothermophilus* GAPDH-coding gene has been inserted in an orientation opposite to that of the Tc-resistance gene. In other words, the high expression observed for this gene should result from the existence in the 5'-non-coding region of the insert of a transcriptional and a translational promoter recognized by the *E. coli* transcriptional and translational machinery.

(1) Ribosome-binding sequence

An RBS, consisting of 7 nt complementary to the 3'-end of *B. stearothermophilus* and *E. coli* 16S rRNAs was observed 9 bp upstream from the ATG

start codon (Fig. 2). The stability of the possible base-pairing is of -11.1 kcal/mole, according to the recent free-energy parameters for RNA duplex stability described by Freier et al. (1986). Such a strong complementarity between mRNA and 16S rRNA is a characteristic of Gram-positive bacteria (Moran et al., 1982; Hager and Rabinowitz, 1985). Fig. 4 represents a comparative study of possible mRNA-16S rRNA hybrids in *B. stearothermophilus*. As can be seen in this figure, the distance between the 16S rRNA 3'-end and the AUG start codon are conserved in the putative duplexes. Indeed the number of nt between the RBS and the AUG codon on the mRNA decreases when the distance between the base-paired region of 16S rRNA and the 16S rRNA 3'-end increases.

The comparison represented in Fig. 4 also reveals that there is no correlation between the stability of the mRNA-16S rRNA hybrid and the amount of the corresponding protein in *B. stearothermophilus* cells. Indeed, the cellular amount of GAPDH is 100 to 200 times superior to that of tRNA-synthetase (Atkinson et al., 1979), whereas tryptophanyl-tRNA synthetase mRNA can form a duplex with 16S rRNA which is theoretically as stable as the one formed with GAPDH mRNA. Other parameters should modulate expression, i.e., transcription rate, as well as mRNA and/or protein stability.

TABLE II
Comparison of codon usage

aa	codon	Codon usage ^a				aa	codon	Codon usage ^b			
		<i>E.</i> GAPDH	<i>B.</i> GAPDH	<i>B.</i> TyrS	<i>B.</i> LDH			<i>E.</i> GAPDH	<i>B.</i> GAPDH	<i>B.</i> TyrS	<i>B.</i> LDH
Asp	GAT	7	7	8	7	Val	GTT	21	7	4	4
	GAC	18	13	15	13		GTC	1	18	10	13
Asn	AAT	1	2	4	8	Met	GTA	8	1	1	3
	AAC	17	18	8	8		GTG	4	14	0	10
Thr	ACT	12	0	0	1	ATG	7	8	6	6	
	ACC	15	5	4	2	Ile	ATT	1	7	13	17
	ACA	0	2	2	1		ATC	19	13	14	10
	ACG	0	12	23	8		ATA	0	0	1	1
Ser	TCT	6	1	2	1	Leu	TTA	1	2	3	4
	TCC	9	2	0	2		TTG	0	19	12	6
	TCA	0	0	2	1		CTT	0	1	10	3
	TCG	0	13	6	3		CTC	0	2	10	5
	AGT	0	0	1	1		CTA	0	0	1	2
	AGC	0	2	7	5		CTG	19	3	7	3
Glu	GAA	13	19	17	9	Tyr	TAT	2	2	4	5
	GAG	2	4	22	10		TAC	6	6	10	7
Gln	CAA	0	3	12	6	Phe	TTT	1	2	12	8
	CAG	5	0	5	1		TTC	10	3	7	6
Pro	CCT	0	0	0	0	Lys	AAA	26	20	18	11
	CCC	0	0	0	0		AAG	1	3	8	3
	CCA	1	0	2	0	His	CAT	1	7	3	6
	CCG	8	11	9	9		CAC	5	2	4	2
Gly	GGT	16	1	0	2	Arg	CGT	8	1	3	1
	GGC	14	16	19	12		CGC	4	12	19	11
	GGA	0	4	4	5		CGA	0	1	1	3
	GGG	0	2	13	9		CGG	0	1	5	3
Ala	GCT	28	4	3	6		AGA	0	0	0	0
	GCC	1	12	9	18		AGG	0	0	1	
	GCA	4	4	3	2	Trp	TGG	3	2	6	3
	GCG	2	19	19	8		stop	TAA	1	1	1
Cys	TGT	0	0	0	0	TAG	0	0	0	0	
	TGC	3	2	2	2	TGA	0	0	0	1	

^a Comparison of the codon usage of *E. coli* (E) (Branlant and Branlant, 1985) and *B. stearothermophilus* (B) (this report) GAPDH and of three different *B. stearothermophilus* enzymes: the GAPDH, the tyrosyl-tRNA synthetase (TyrS; Winter et al., 1983) and the lactate dehydrogenase (LDH; Barstow et al., 1986b).

^b Numbers indicate how many times each codon is used in the gene coding for the protein.

TABLE III

Comparison of G + C content

Gene ^a	% G + C	% G + C at third nt of the codon	Reference
<i>gapD</i>	50.4	51.8	Branlant and Branlant (1985)
Bst <i>gap</i>	55.1	71.1	This report
Bst <i>trpS</i>	54	69	Barstow et al. (1986a)
Bst <i>ldh</i>	52	60	Barstow et al. (1986b)
Bst <i>pfk</i>	58.5	71.3	French and Chang (1987)

^a Comparison of the G + C content of *E. coli gapD*, *B. stearothersophilus* GAPDH (Bst *gap*), *B. stearothersophilus* tryptophanyl-tRNA synthetase (Bst *trpS*) and *B. stearothersophilus* lactate dehydrogenase (Bst *ldh*), *B. stearothersophilus* phosphofructokinase (Bst *pfk*).

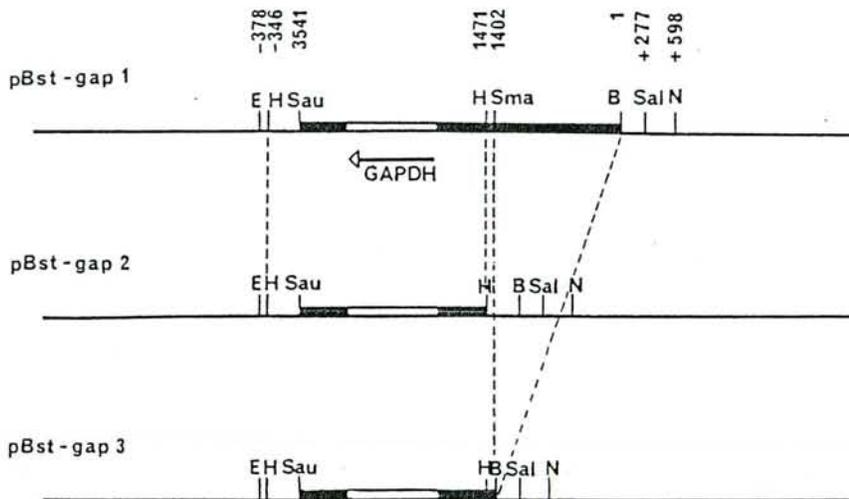


Fig. 3. Construction of plasmids pBst-gap2 and pBst-gap3.

Plasmid pBst-gap2 was obtained by insertion of the *Hind*III fragment (nt -346 to 1471) at the *Hind*III site of pBR322. The following experimental procedure was used to construct plasmid pBst-gap3: plasmid pBst-gap1 was digested with *Bam*HI + *Sma*I enzymes. After phenolic extraction and alcohol precipitation, the recessed 3'-ends of the double-stranded DNA fragments were filled-in in the presence of dNTPS using the Klenow fragment of *E. coli* DNA polymerase I, as described by Maniatis et al. (1982). The mixture was then ligated with T4 DNA ligase and used to transform the strain DF221 *gapD*⁻ (kindly provided by Prof. D.G. Fraenkel) as previously described (Branlant et al., 1983). Elimination of the *Sma*I-*Bam*HI fragment in the resulting plasmid was verified by digestion with *Hind*III and *Sal*I ends nucleases. The heavy line represents the insert: open area = coding region; filled-in line = flanking regions, the thin line represents the pBR322 vector. The unique restriction sites used in the experiments are indicated. B: *Bam*HI, E: *Eco*RI, H: *Hind*III, Sal: *Sal*I, Sau: *Sau*3A, Sma: *Sma*I, N: *Nru*I. Numbering in Fig. 3 is according to numbering in Figs. 1 and 2.

(2) Transcriptional promoter

No typical promoter-like structure was found in close vicinity to the RBS region, as for instance in the case of the gene coding for L-lactate dehydrogenase (Barstow, 1986b).

In order to determine the approximate location of the promoter region, two new plasmids pBst-gap2 and pBst-gap3, were constructed (Fig. 3) and were tested for the production of *B. stearothersophilus* GAPDH in the *E. coli gapD*⁻ mutant strain DF221.

Plasmid pBst-gap2 was obtained by insertion of the *Hind*III fragment of pBst-gap1 containing the coding region into the *Hind*III site of pBR322. In this way, the size of the *B. stearothersophilus* DNA region flanking the 5'-end of the coding region in the new plasmid (pBst-gap2) was reduced to 564 nt. Plasmid pBst-gap3 was obtained by elimination of the *Sma*I-*Bam*HI fragment of the insert, so that in the new plasmid (pBst-gap3) the upstream region was reduced to 633 nt.

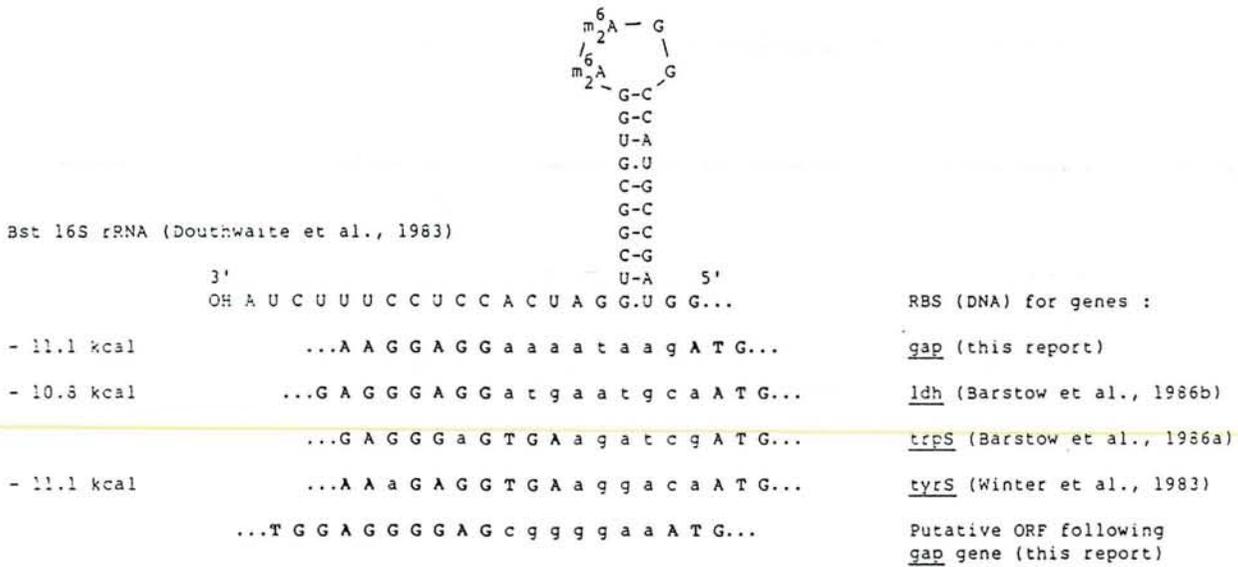


Fig. 4. Comparison of the sequences of ribosome-binding sites (RBS) from different *B. stearothermophilus* genes.

The *B. stearothermophilus* 16S rRNA 3'-terminal structure is from Douthwaite et al. (1983). For RBS (DNA), the start codons and the sequences complementary to 16S rRNA are represented by capital letters. Estimations of base-pairing stability are according to Freier et al. (1986). The putative RBS of the ORF following the GAPDH-coding gene is indicated.

None of these plasmids were able to complement strain DF221 for GAPDH. Indeed, strain DF221 transformed with pBst-gap2 or with pBst-gap3 was incapable of growing on glucose/Ap M63 medium. This result indicates that the promoter site is located upstream of the *Sma*I site situated 633 bp upstream from the ATG start codon.

Computer analysis of the 5'-non-coding region of 2034 nt reveals a unique strong promoter-like structure with -10 and -35 consensus sequences of *E. coli* promoters. It is located between nt coordinates 853 and 882 (Fig. 2), that is, 1153 nt upstream from the 5'-end of the coding region. Another sequence with a strong homology to the -10 region of *E. coli* promoter is present at nt position 9 of the insert. Transcription from this 'TATA' box would require the presence of a -35-like sequence at the right position in the pBR322 vector; a sequence with poor homology to the -35 region is found at the expected position. Transcription of the cloned gene in *E. coli* probably starts at the promoter-like region between nt 853 and 882. Further experiments are required to verify this assumption.

The fact that the promoter site is quite distant from the coding region probably explain why only large *B. stearothermophilus* chromosomal DNA segments were found to complement the *E. coli gapD*⁻ mutant

(Branlant et al., 1983), in contrast to the situation observed for *E. coli*, where a chromosomal DNA fragment of 1523 bp was able to complement this mutant.

No obvious ORF was seen in the DNA region situated between the putative promoter and the 5'-end of the GAPDH coding region. And indeed no other protein was found to be highly expressed together with GAPDH when strain DF221 was transformed with plasmid pBst-gap1 (Branlant et al., 1983). Therefore, the distance between the putative promoter region and the coding region remains puzzling. Thus it is clear that, in regard to transcriptional promoter of GAPDH-coding genes quite different situations are observed in *E. coli* (two strong overlapping promoter-like sequences close to the ATG start codon) and in *B. stearothermophilus*.

(3) Transcriptional terminator

A putative Rho-independent termination signal with a G + C-rich hairpin structure followed by a stretch of T residues is found 57 bp downstream from the TAA termination codon (Fig. 5). Here again it is a relatively long distance as compared to usual termination sites of *E. coli*. Interestingly, this termination-like signal is preceded by a 'TATA' box sequence. An ATG is found at the top of the hairpin

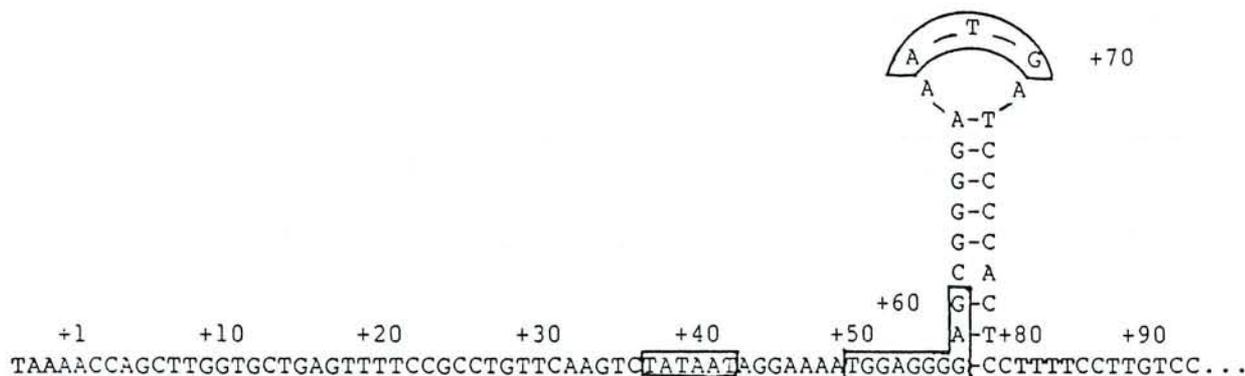


Fig. 5. Structure of the putative Rho-independent transcription termination signal following *B. stearothersophilus* GAPDH-coding gene. Nucleotides have been numbered starting from the TAA termination codon. The putative 'TATA' box, RBS and initiation codon of the ORF following GAPDH-coding gene are boxed.

loop, preceded by a long putative RBS (Figs. 4 and 5). No characteristic -35 box is observed before the putative 'TATA' box (Fig. 2). Nevertheless, since the putative ATG start codon is followed by an ORF extending up to the 3'-end of the 3541-bp DNA fragment, this may represent the beginning of a gene coregulated with the GAPDH-coding gene.

(c) Conclusions

(1) The correction made in this paper to be previously published amino acid sequence of *B. stearothersophilus* GAPDH has important consequences for the interpretation of x-ray diffraction data obtained from crystals of this enzyme.

(2) Codon utilization of the *B. stearothersophilus* GAPDH-coding gene is characterized by a marked preference for G or C residues at the third position of codons. Although *B. stearothersophilus* GAPDH-coding gene contains several codons rarely found in *E. coli*, it is highly expressed in *E. coli*.

(3) The GAPDH-coding region is preceded by a strong RBS characteristic of Gram-positive bacteria.

(4) The putative transcriptional promoter of the GAPDH-coding gene is situated at a long distance from the ATG initiation codon (1153 nt).

(5) The GAPDH-coding gene is followed by a putative Rho-independent termination signal, which precedes an ORF. The transcription and translation signals corresponding to this ORF overlap the putative Rho-independent termination signal.

ACKNOWLEDGEMENTS

This work was supported by grant No. 84V0670 of the Ministère de la Recherche et de la Technologie and the Centre National de la Recherche Scientifique: Essor des Biotechnologies.

ADDENDUM

After we had submitted this paper, Conway and Ingram (1988) published the cloning and sequencing of the phosphoglycerate kinase (PGK) gene from *Zymomonas mobilis*. They found that in this Gram-positive organism, the phosphoglycerate-kinase-coding gene belongs to the *gap* operon. We looked for a possible homology between the peptide encoded by the N-terminal fragment of unknown gene just following GAPDH in the sequenced 3541-bp *B. stearothersophilus* DNA fragment and known primary structures of phosphoglycerate kinase. We concluded that this gene fragment should indeed code for PGK N-terminal segment.

REFERENCES

- Atkinson, T., Banks, G.T., Bruton, C.J., Comer, M.J., Jakes, R., Kamalagharan, T., Whitaker, A.R. and Winter, G.P.: Large scale isolation of enzymes from *Bacillus stearothersophilus*. *J. Appl. Biochem.* 1 (1979) 247-258.

- Barstow, D.A., Sharman, A.F., Atkinson, T. and Minton, N.P.: Cloning and complete nucleotide sequence of the *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase gene. *Gene* 46 (1986a) 37-45.
- Barstow, D.A., Clarke, A.R., Chia, W.N., Wigley, D., Sharman, A.F., Holbrook, J.J., Atkinson, T. and Minton, N.P.: Cloning, expression and complete nucleotide sequence of the *Bacillus stearothermophilus* L-lactate dehydrogenase gene. *Gene* 46 (1986b) 47-55.
- Biesecker, G., Harris, J.I., Thierry, J.C., Walker, J.E. and Wonacott, A.J.: Sequence and structure of D-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *Nature* 266 (1977) 328-333.
- Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F.: Codon-defined ribosomal pausing in *Escherichia coli* detected by using the *pyrE* attenuator to probe the coupling between transcription and translation. *Nucleic Acids Res.* 13 (1985) 4113-4123.
- Branlant, G. and Branlant, C.: Nucleotide sequence of the *Escherichia coli* *gap* gene. *Eur. J. Biochem.* 150 (1985) 61-66.
- Branlant, G., Flesch, G. and Branlant, C.: Molecular cloning of the glyceraldehyde-3-phosphate dehydrogenase genes of *Bacillus stearothermophilus* and *Escherichia coli* and their expression in *Escherichia coli*. *Gene* 25 (1983) 1-7.
- Conway, T. and Ingram, L.O.: Phosphoglycerate kinase gene from *Zymomonas mobilis*: cloning, sequencing, and localization within the *gap* operon. *J. Bacteriol.* 170 (1988) 1926-1933.
- Douthwaite, S., Christensen, A. and Garrett, R.A.: Higher order structure in the 3'-minor domain of small subunit ribosomal RNAs from a Gram negative bacterium, a Gram positive bacterium and an eukaryote. *J. Mol. Biol.* 169 (1983) 249-279.
- Freier, S.M., Kierzek, R., Jaeger, J., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H.: Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* 83 (1986) 9373-9377.
- French, A. and Chang, S.: Nucleotide sequence of the phosphofructokinase gene from *Bacillus stearothermophilus* and comparison with the homologous *Escherichia coli* gene. *Gene* 54 (1987) 65-71.
- Hager, P.W. and Rabinowitz, J.C.: Translational specificity in *Bacillus subtilis*. In Dubnau, D.A. (Ed.), *The Molecular Biology of the Bacilli*, Vol. II. Academic Press, New York, 1985, pp. 1-32.
- Leslie, A.G. and Wonacott, A.J.: Coenzyme binding in crystals of glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.* 165 (1983) 365-391.
- Leslie, A.G. and Wonacott, A.J.: Structural evidence for ligand-induced sequential conformational changes in glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.* 178 (1984) 743-772.
- Maniatis, T., Fritsch, E.F. and Sambrook, J.: *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Moran, C.P., Land, N., Le Grice, S.F., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R.: Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* 186 (1982) 161-178.
- Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarrantor, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G.: Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res.* 12 (1984) 6663-6671.
- Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74 (1977) 5463-5467.
- Sharp, P.M., Wen-Hsiung Li: Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.* 14 (1986) 7737-7749.
- Skarzynski, T., Moody, P.C.E. and Wonacott, A.J.: Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.* 193 (1987) 171-187.
- Varenne, S. and Lazdunski, C.: Effects of distribution of unfavorable codons on the maximum rate of gene expression by an heterologous organism. *J. Theor. Biol.* 120 (1986) 99-110.
- Walker, J.E., Wonacott, A.J. and Harris, J.I.: Heat stability of a tetrameric enzyme, D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* 108 (1980) 581-586.
- Winter, G.P., Koch, G.L., Hartley, B.S. and Barker, D.G.: The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. *Eur. J. Biochem.* 132 (1983) 383-387.

Communicated by J.P. Lecocq.

Complément de la publication

Il persistait une incertitude quant au codon d'initiation du gène de la PGK de *B.stearothermophilus*. En effet, lors de l'analyse de la seconde phase ouverte, nous avons observé que la structure primaire déduite présentait de fortes homologues avec les autres séquences déterminées de PGK seulement à partir du 19^e résidu amino-acide qui est une méthionine. L'ATG correspondant, en position 3.164, pourrait être le codon d'initiation réel du gène de la PGK. Ceci est d'autant plus probable qu'il est précédé d'une séquence présentant une grande complémentarité avec l'extrémité 3' du rRNA 16S de *B.stearothermophilus* (Douthwaite et al., 1983). Cette séquence de fixation du ribosome est constituée des nucléotides 3.147 à 3.155 et est localisée à 9 nucléotides en amont de l'ATG (Fig. 47).

La comparaison que nous avons faite dans la publication précédente, au niveau des séquences de fixation du ribosome (Fig. 4 de la publication), montrent un positionnement très strict de l'AUG initiateur par rapport à l'extrémité 3' du rRNA 16S (1 nucléotide maximum de décalage). L'AUG en position 19 et le site correspondant de fixation du ribosome répondent à ce critère (Fig. 2 de la publication). L'AUG en position 1 et le site correspondant de fixation du ribosome ne répondent pas à ce critère : l'AUG est décalé de 2 nucléotides par rapport à celui du gène de la GAPDH et de 3 nucléotides par rapport à ceux de la lactate déshydrogénase et des tryptophanyl et tyrosyl-tRNA synthétases. D'autre part, la séquence possible de fixation du ribosome s'étend inhabituellement jusqu'à l'extrémité 3' du rRNA 16S, les 3 derniers nucléotides de cette séquence de fixation du ribosome étant de plus impliqués dans un appariement de bases, précisément dans la structure de terminaison de la transcription (Fig. 5 de la publication).

Il est vraisemblable, pour toutes les raisons que nous venons de citer, que l'efficacité de fixation du ribosome au niveau du site correspondant à l'AUG 19 soit très supérieure à celle de la fixation au niveau du site correspondant à l'AUG 1.

Néanmoins, il n'est pas exclu que la présence de l'AUG 1, précisément au sommet de la boucle de la structure de terminaison, et du site correspondant de fixation du ribosome, au niveau du "stem" de cette structure, soit liée à un processus de co-régulation de l'expression des deux gènes. La fixation d'un ribosome au niveau de cette structure pourrait peut-être empêcher l'arrêt de la transcription par le signal de terminaison. Ce point est à rapprocher du fait que le signal de terminaison est bien précédé d'un signal possible d'initiation de la transcription pouvant constituer un signal de réinitiation propre au gène de la PGK.

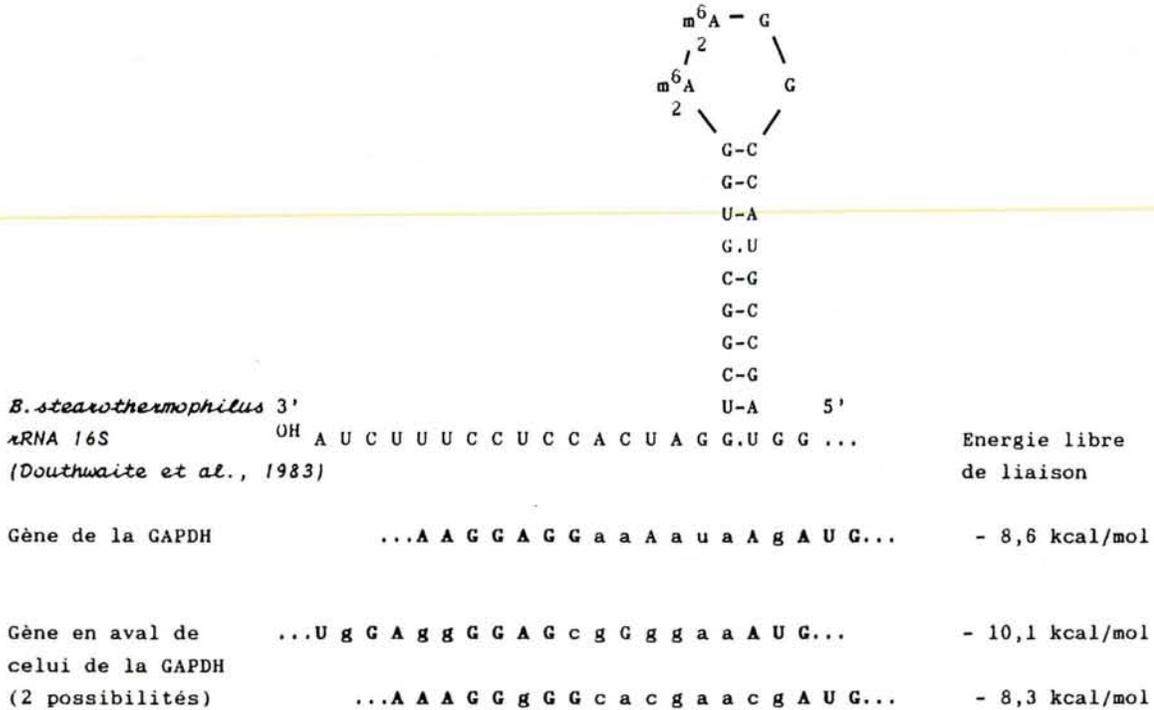


Figure 47 : Les séquences RBS potentielles des gènes de *B. stearothermophilus* contenus dans pBst-gap1

Les séquences RBS sont indiquées en gras, ainsi que les codons d'initiation de la traduction.
 Les nucléotides s'appariant à la séquence du rRNA 16S sont inscrits en majuscules.
 (L'énergie libre de liaison avec l'extrémité 3' du rRNA 16S est calculée selon Freier et al., 1986.)

Mais, alors que la séquence -10 de ce signal répond parfaitement au consensus des séquences -10 aussi bien d'*E.coli* que de bactéries Gram⁺, une séquence présentant une faible homologie avec les séquences -35 de bactéries Gram⁺ est trouvée à 23 nucléotides en amont de la séquence -10. Cette distance est extrêmement importante par rapport à ce qui a été observé dans l'ensemble des gènes de Gram⁺ séquencés jusqu'à ce jour (Graves et Rabinowitz, 1986). Il est donc possible que ce signal d'initiation de la transcription soit faible, ou bien qu'il soit soumis à une régulation très stricte.

Une étude plus approfondie de transcription et de transcription-traduction *in vitro* devra être effectuée afin de vérifier ces différents points.

IV - ETUDE DE L'EXPRESSION DU GENE DE LA GAPDH A PARTIR DU PLASMIDE pBst-gap4

Un autre problème a été soulevé lors de l'étude du plasmide pBst-gap4. La seule séquence susceptible de constituer un promoteur transcriptionnel efficace du gène de la GAPDH est située dans la région 5' du fragment inséré dans pBst-gap1, absente dans pBst-gap4.

L'expression du gène de la GAPDH inséré dans ce plasmide pBst-gap4 a été étudiée dans la souche DF221. L'activité enzymatique, mesurée dans l'extrait brut obtenu par sonication, équivaut à 1/15^e environ de celle dosée dans l'extrait brut des cellules DF221 transformées par le plasmide pBst-gap1. Le gène de ce plasmide peut donc s'exprimer à partir d'un autre promoteur que celui identifié dans pBst-gap1.

Cet autre promoteur aurait pu appartenir au plasmide pBR322, tel que le promoteur bla, présent dans pBst-gap4 et orienté dans le même sens que le gène de la GAPDH duquel il est cependant très éloigné. Mais le gène est orienté dans le même sens également dans les plasmides pBst-gap2 et pBst-gap3 décrits dans la figure 3 de la publication. Ces plasmides ne complètent pas la souche DF221, donc ne produisent pas de GAPDH. La différence entre ces plasmides et pBst-gap1 est l'absence dans pBst-gap2 du fragment BamHI - HindIII (nucléotides 1 à 1.471) et dans pBst-gap3 du fragment BamHI - SmaI (nucléotides 1 à 1.402).

L'hypothèse de l'expression du gène à partir du promoteur bla était donc à exclure, ce qui est normal puisque le gène bla est suivi d'un signal de terminaison de la transcription. Le taux de transcrits auquel nous aurions pu nous attendre aurait été extrêmement faible.

Restaient deux hypothèses possibles. La première suggérait la présence d'un très petit fragment en 5' du fragment inséré dans pBst-gap4 et absent dans pBst-gap1. Il faudrait que ce fragment soit indécélable à la fois sur les produits d'hydrolyse par HaeIII et Sau3A. Cela remettrait en cause la colinéarité de la région 1 - 974 avec la région qui suit. Cette hypothèse était très peu vraisemblable.

La deuxième hypothèse était l'existence d'un promoteur de faible efficacité, entre les nucléotides 975 et 1.402, pouvant expliquer la faible expression du gène à partir de pBst-gap4. Nous avons donc recherché d'éventuelles séquences promotrices entre les nucléotides 975 et 1.402. Nous avons observé deux signaux possibles, en positions 1.026 - 1.063 et 1.193 - 1.229 (Fig. 48). Ces signaux se caractérisent par des séquences -10 n'ayant que le TA en 5' et le T en 3' conservés par rapport au consensus et des séquences -35 présentant la séquence TTGA. Cependant, dans chaque signal, ces séquences -10 et -35 sont très éloignées l'une de l'autre (24 et 23 nucléotides), beaucoup plus que ce qui est normalement observé (17 à 18 nucléotides). Il est possible néanmoins que ces deux signaux puissent contribuer à l'expression du gène de la GAPDH dans pBst-gap4, aucun signal de ce type n'étant présent dans la région 1.402 - 2.034 précédant la région codante.

Des expériences de transcription *in vitro* seront nécessaires pour étudier l'efficacité relative des différents promoteurs possibles. Néanmoins, nous pouvons d'ores et déjà conclure que, dans le cas de *B.stearothermophilus*, le promoteur en amont de l'opéron est situé à au moins 632 nucléotides en amont du codon d'initiation du gène de la GAPDH.

V - LA STRUCTURE EN TANDEM DES GENES DE LA GAPDH ET DE LA PGK

Nous venons de voir que, tout comme chez *C.pasteurianum* ou *Z.mobilis* (Conway et Ingram, 1988), les gènes de la GAPDH et de la PGK sont organisés en tandem chez *B.stearothermophilus*. Comme nous l'avons déjà mentionné, cette organisation n'a pas été retrouvée chez *E.coli*.

La structure en opéron des gènes de la GAPDH et de la PGK, observée chez chacun des trois micro-organismes Gram+ étudiés, est vraisemblablement générale à l'ensemble des bactéries Gram+. Une telle organisation peut permettre une co-régulation étroite des deux enzymes. Ceci est à rapprocher du fait que, selon Ashmarina et al. (1985), ces deux enzymes pourraient être associées *in vivo* en un complexe enzymatique.

	GATCCTGCT	9													
ATAATAGAACATGCAACGGAACGAAATAGAGAAATCATGAACTTCCATGCCCTCGTAGTGTAGTGGATAGCACGA		84													
GAGATTCCGGTTCTCTTAGCGTGGGTTTCAATCCTGCCGAGGGCGTTCCAACGATCTGACAAGTTCGGATTGAGA		159													
TTCCCAACCCCTTGATACGACCGCGTTTGCGGTCTTTTTTCATTGCTTGCCGAGCTGGCGGCGCGTTTTCCGC		234													
ATCGGAACAAGAGATGGAACGGGGGCTGCGCCTTGTGCAGTCGCTCGACCCGCCGGGAGTCGGCGCCCGGATT		309													
GGCGGAATGCCTCTTGCTGCAGCTTGAACGTCTTCTGAGCGCGACGAGCTGGCGGAAACGATCGAGGTGGCTTT		384													
CCTCCGCTCAAGTAAAAGCGATCATCCAAGAGCTGATCGCGGCGAAGACCGACAGACGCGCTTTCTGGCCAAC		459													
AGCTTGCTGATTGCTTTCATGGCCGATACGGCATCGCCATTTGCTCCGCACGCGTCCGCAATACCGCGAGCAG		534													
CTGGCATTCTTCTTTCGGCGAAACGAAAGCAATACACCGGAGCTTGATGCGACCGGTGAAAGAAAGGAAGAACG		609													
CGATGCACATCCGGTTGTACACGAAGACGAACTGTCCGCTTTGCGCAAAAGCGAAAGCGTTTTGACGGAGCTGC		684													
AGGCAGACTATCGGTTTACGATGGAAGAGATCGATATTTATGAAGACGACGCGCTTCTCGAGAAATATCAGTTGA		759													
TGATTCCGGTCGTCGAGCTTGACGGGGAAGAAATCGGGTATGGGATGATTGAAAAAGAAACGGTGAGAAAACGGT		834													
TGCGGCGGGCAAAAATA	[]	GTTGAAC	[]	TGCTATCGTCTCTCTGC	[]	TATAAT	[]	GAAAATAGCGGGGAGGAATTTT	905						
TTATCGATTGGCGGGACATAATATGTCATACCGGGACAAAAACGTCCTATGTTCTGCTAAGGGAAAGGATCGA									980						
TGCAACCGTTAGTAGAGGCATTAAAAAAATTATCGCCCGACTTGC						[]		TTGACGTT	[]	ATGCAAAAACGATATCAAAT	1053				
TTTG	[]	TACTCT	[]	ATTCGCTTATGGCGCCGATCGGGCGGCGGCGCTGGCGGCGAGCCTCGGCATGAGCGAGCGGG						1126					
TTCTCCGGTCGGAACAGAATTTTTGAAAGGACAAAACCTTGCTTTCCACCGATGTGTGCGGGCATGCG								[]	GTTGACG	1200					
GAGGAAGGCCAGGCGCTCCTTCA				[]	TACGCT	[]	CAATGATTTGATGCGGGAAGCGCTCGGACTCAAAGAGCTGGAAG			1273					
CCGCGCTGCAGGAACAGCTCGGCATTCCGCGGGTGATTGTGCTTGCCGGGGATAGCGACCGTCCCCTTGGGTAA										1348					
AAAAAGAAATGGGTAGAGCGTGTGTGCGCTGCATGAAGGAGCTGCTTGAGCCCGGGACATCGTCGCGTGGCCGG										1423					
CGGAACGACGATGGCGGCGGTGCGGGAGATGATGACGCCGATTGAAAGCTTCGTGATGTGTGTTGTTGTCGCCGC										1498					
GCGCGCGGGCCTTGGCGAAGATGTCGAGAATCAGGCGAACACGATTTGTGCGAAAATGGCGGAAAAGCGATGGG										1573					
CCGCTATCGGTTGCTGCACGTGCGCTGACCAGCTGAGCGATGAAGCGTACGCGTCGCTGTGCGAAGAGCCGGCGGT										1648					
GAAAGAAGTGCTTGAGCTCATCCAATCATGCCGCATGGTCTCCATGGCATCGGTGAAAGCGGTGACGATGGCCAA										1723					
GCGGCGGAAGACGCGCCTGTAGAGATGGAGAACATCATCGCCCGCCATGGCCGCGGAGGCTTCGGC								[]	TACTAT	1796					
TTCAACGAGTACGGCGACGTCGTCCTATAAAGTAAAACAGTCGGCATTCAACTTGAACACCTCCCGCATGTGCGAG										1871					
CATGTATCGCCGTCGCGGAGGGCGTCAAGGCGAAAGCATTGCGCGGTACATGAAACGGGCGCCGACTCGAT										1946					
ATTGGTGACGGACGAAGGAGCCGCAAAAGCGTTCGTAGGGGAGTAATCTCCCTTGCTGATACAATCCAATCGCAA										2021					
GGAGGAAAATAAG ATG GCA GTC AAA GTG GGA ATC AAC GGA TTT GGC CGC ATC GGA CGC										2079					
	M	A	V	K	V	G	I	N	G	F	G	R	I	G	R
	1												10		

Figure 48 : Localisation des différentes séquences promotrices potentielles en amont du gène de la GAPDH de *B.stearothermophilus*

1. Le promoteur transcriptionnel en amont de l'opéron

Les gènes de la GAPDH de *C.pasteurianum*, de *Z.mobilis*, tout comme le gène d'*E.coli*, présentent la particularité de posséder deux signaux possibles d'initiation de la transcription présentant un recouvrement de séquence.

B.stearothermophilus représente certainement un cas particulier en raison de l'éloignement très important du promoteur transcriptionnel. Deux promoteurs putatifs, sans doute de faible efficacité, ont pu être identifiés dans un segment colinéaire à la partie codante. Un promoteur de forte efficacité a été localisé à 1.153 nucléotides de l'ATG, dans un segment vraisemblablement colinéaire au gène, d'après nos résultats. Une vérification reste cependant nécessaire à ce niveau.

2. La région intergénique (GAPDH - PGK)

2-1. Les signaux de terminaison de la transcription

Chez *C.pasteurianum* comme chez *Z.mobilis*, le gène de la GAPDH est suivi de séquences répétées inversées susceptibles de conduire à des structures secondaires de faibles énergie libre au niveau du mRNA. Ces structures conduisent peut-être à des pauses de la RNA polymérase DNA-dépendante, voire à des décrochements occasionnels. Dans chaque cas, cette structure est localisée à une distance très faible du codon de terminaison.

Le gène de la GAPDH de *B.stearothermophilus* est suivi lui d'un signal de terminaison Rho-indépendant riche en G + C, correspondant à une structure secondaire très stable au niveau du mRNA. Curieusement, cette structure présente un ATG au sommet de la boucle et une séquence Shine-Dalgarno dans l'appariement.

Ces signaux ne semblent pas être ceux correspondant réellement au gène de la PGK, mais ils pourraient intervenir dans un processus de co-régulation des deux gènes. Là encore, *B.stearothermophilus* diffère donc des deux autres bactéries Gram⁺.

2-2. Les signaux d'initiation de la transcription et de la traduction

Les régions codantes des gènes de la PGK de *C.pasteurianum* et de *B.stearothermophilus* sont précédées de séquences -10 promotrices potentielles pouvant permettre une réinitiation de la transcription, le signal comprenant également une séquence -35 chez *C.pasteurianum*.

Dans ce cas, les productions de GAPDH et de PGK pourraient être régulées différemment lors de l'étape transcriptionnelle. Chez *B.stearothermophilus*, la séquence -10 du promoteur de la PGK est localisée à 79 nucléotides en amont du codon ATG situé en position 3.164 qui semble être le véritable codon d'initiation. La séquence -10 se situe à 15 nucléotides en amont de la structure terminatrice en aval du gène de la GAPDH (Fig. 5 de la publication), la séquence -35 étant très éloignée et présentant une faible homologie avec la séquence consensus. Cette organisation particulière intervient peut-être dans la co-régulation de l'expression de ces deux gènes.

Chez *Z.mobilis*, Conway et Ingram (1988) ont observé que les activités spécifiques de la GAPDH et de la PGK n'étaient pas couplées. L'absence de séquences promotrices de la transcription entre les deux gènes montre que cette différence n'est pas liée à un contrôle transcriptionnel. Les auteurs ont pensé à une sensibilité différente de chacune des deux enzymes au pH intracellulaire. La présence de deux sites potentiels de fixation du ribosome et au moins trois séquences palindromiques susceptibles de constituer des structures secondaires au niveau du mRNA les a conduits à suggérer un contrôle lors de l'étape de la traduction.

Chez chacune des bactéries étudiées, la partie codante du gène de la PGK est précédée d'un signal potentiel d'initiation de la traduction.

3. Etude comparative de l'expression des gènes dans *E.coli*

Chacun des gènes de GAPDH clonés dans *E.coli* peut s'exprimer à partir de son propre promoteur. Le taux de cette expression est variable selon la bactérie d'origine. En effet, le gène d'*E.coli* est très fortement surexprimé chez *E.coli*. Celui de *B.stearothermophilus* est lui aussi fortement exprimé à partir du promoteur localisé à 1.153 nucléotides en amont. Il faut souligner que, dans le cas de *B.stearothermophilus*, cette surexpression est obtenue malgré un usage de codons assez défavorable chez *E.coli* (Tab. II et III de la publication). Ceci indique donc qu'un promoteur de *B.stearothermophilus* peut conduire à une expression très forte chez *E.coli*.

Par contre, l'expression des gènes de la GAPDH de *C.pasteurianum* et de *Z.mobilis* à partir de leur propre promoteur est nettement plus faible. Dans chaque cas, une expression plus forte a pu être obtenue en plaçant ces gènes sous le contrôle d'un autre promoteur tel que le promoteur lacZ. La surexpression obtenue restait néanmoins plus faible que celle observée dans le cas de *B.stearothermophilus* et d'*E.coli*. Ceci peut s'expliquer par l'usage particulièrement défavorable des codons dans le cas du gène clostridial. Cette explication n'est pas valable pour le gène de *Z.mobilis* dont l'usage des codons est voisin de celui des gènes fortement exprimés chez *E.coli*, *Z.mobilis* ayant un usage préférentiel des codons similaire à celui d'*E.coli* sauf pour Ala et Tyr.

VI - CONCLUSIONS

Cette étude sur *B.stearothermophilus* a donc permis d'apporter un argument supplémentaire en faveur de l'organisation générale en opéron des gènes de la GAPDH et de la PGK chez les bactéries Gram⁺.

La forte expression du gène de la GAPDH de *B.stearothermophilus* chez *E.coli*, malgré la composition en bases riche en G + C, caractéristique du génome de *B.stearothermophilus*, et l'utilisation de codons correspondant à des tRNA mineurs chez *E.coli*, montre qu'un promoteur de *B.stearothermophilus* peut être très efficace chez *E.coli*.

Cette forte expression et la séquence nucléotidique que nous avons déterminée ont permis de développer au laboratoire une étude des relations structure-fonction de l'enzyme par mutagenèse dirigée.

La structure primaire de la GAPDH de *B.stearothermophilus* est à présent déterminée sans ambiguïté. Nous l'avons utilisée dans l'étude comparative des GAPDH de différentes espèces que nous décrivons dans le chapitre V.

Chapitre V :

Etude phylogénétique
de la GAPDH

SOMMAIRE

I - ALIGNEMENT DES SEQUENCES DE GAPDH.....	page 193
II - REGIONS CONSERVEES ET REGIONS DIVERGENTES DE LA GAPDH.....	page 194
1. Domaine de fixation du coenzyme.....	page 195
1-1. Les régions fortement conservées.....	page 195
1-2. Les régions fortement divergentes.....	page 200
1-3. Conclusion.....	page 200
2. Domaine catalytique.....	page 201
2-1. Les régions conservées.....	page 201
2-2. Les régions variables.....	page 203
2-3. Conclusion.....	page 203
III - LA FILIATION DES ESPECES D'APRES L'ETUDE DE LA GAPDH.....	page 204
1. Comparaisons des GAPDH bactériennes.....	page 204
1-1. Comparaison entre <i>E.coli</i> et les quatre bactéries Gram+.....	page 204
1-2. Comparaison entre les GAPDH de bactéries Gram+ mésophiles et thermophiles.....	page 204
1-3. Conclusion.....	page 206
2. Comparaisons entre les GAPDH bactériennes et les GAPDH chloroplastiques.....	page 209
3. Comparaisons entre les GAPDH bactériennes et les GAPDH cytosoliques d'eucaryotes.....	page 212
IV - CONCLUSIONS.....	page 213

Nos travaux ont permis de déterminer la structure primaire des GAPDH de *C.pasteurianum* et de *B.stearothermophilus*. Nous disposons donc de la séquence protéique de cinq enzymes bactériennes, celle d'*E.coli* (Branlant et Branlant, 1985), bactérie Gram-, et celles de quatre autres bactéries Gram+ :

- deux bactéries thermophiles : *B.stearothermophilus* (chapitre IV) et *T.aquaticus* (Hocking et Harris, 1980) ;

- deux bactéries mésophiles : *C.pasteurianum* (chapitre II) et *Z.mobilis* (Conway et al., 1987), toutes deux bactéries anaérobies strictes.

Comme nous l'avions déjà souligné dans la partie "Introduction", les séquences de GAPDH de nombreux eucaryotes étaient connues, ainsi que celles de GAPDH chloroplastiques. Une étude phylogénétique pouvait donc être abordée.

Des comparaisons de séquences entre les GAPDH avaient déjà été réalisées, notamment par Martin et Cerff (1986) et Shih et al. (1986), mais la connaissance de nouvelles séquences nécessitait une remise à jour de ces comparaisons. A ce propos, il faut préciser que nos résultats ont été obtenus en considérant que la région C-terminale à partir du résidu 312 appartient au domaine de fixation du coenzyme, conformément à la structure tridimensionnelle déterminée à partir de la GAPDH de *B.stearothermophilus* par Walker et al. (1980) et représentée sur la figure 4 dans la partie "Introduction". D'autres travaux récents utilisent également cette définition du domaine de fixation du cofacteur (Skarzynski et Wonacott, 1988). Cette distinction peut expliquer de légères différences entre nos pourcentages d'homologie et ceux parus dans la littérature.

I - ALIGNEMENT DES SEQUENCES DE GAPDH

L'alignement des différentes séquences a été réalisé de façon à obtenir un maximum d'homologie entre les GAPDH des différents organismes. Les résidus ont été numérotés selon Biesecker et al. (1977).

Nous avons aligné les séquences des enzymes suivantes :

- 5 enzymes bactériennes, citées plus haut ;
- 2 enzymes de levures : *S.cerevisiae* (Holland et al., 1979) et *Z.rouxii* (Imura et al., 1987) ;
- 7 enzymes animales : 2 d'invertébrés, la drosophile (Tso et al., 1985b) et le nématode (Yarbrough et al., 1987) ; 5 d'eucaryotes supérieurs, le homard (Harris et Waters, 1976), le poulet (Dugaiczik et al., 1983), le rat (Tso et al., 1985a), le porc (Harris et Perham, 1968) et l'homme (Tso et al., 1985a) ;
- 6 enzymes végétales : 3 cytosoliques, de la moutarde (Martin et Cerff, 1986), du maïs (Brinkmann et al., 1987) et du tabac (Shih et al., 1986) ; 3 chloroplastiques, une du maïs (Brinkmann et al., 1987) et deux du tabac (Shih et al., 1986).

Nous n'avons pas pris en compte l'enzyme de *T.brucei* et l'enzyme chloroplastique de la moutarde, dont les structures primaires n'ont été que partiellement déterminées (Misset et al., 1987 ; Martin et Cerff, 1986), de même que l'enzyme récemment décrite comme étant la GAPDH de *M.fervidus*, cette enzyme présentant une très faible homologie avec l'ensemble des enzymes des autres organismes (Fabry et Hensel, 1988).

L'alignement réalisé est représenté par la figure 49.

Dans un premier temps, nous effectuerons une étude détaillée des régions de la GAPDH conservées ou divergentes au cours de l'évolution. Nous examinerons la localisation de ces régions dans la structure tertiaire et leur implication dans la fixation du coenzyme ou du substrat ou dans le mécanisme enzymatique.

Nous procéderons ensuite à une étude phylogénétique des espèces en utilisant la GAPDH comme base.

II - REGIONS CONSERVEES ET REGIONS DIVERGENTES DE LA GAPDH

L'alignement que nous proposons (Fig. 49) met en évidence un certain nombre de régions et de résidus conservés dans toutes les séquences de GAPDH. Nous examinerons successivement les régions appartenant au domaine de fixation du coenzyme et celles appartenant au domaine catalytique.

Etudions progressivement la conservation structurale de l'enzyme.

1. Etude des séquences du domaine de fixation du coenzyme

1-1. Les régions fortement conservées

Au niveau de la région N-terminale de l'enzyme, le segment comprenant les résidus de 6 à 13 présente une séquence conservée à 100 % à travers toutes les espèces étudiées. Ce segment intégralement conservé correspond à une boucle et à la portion N-terminale de l'hélice α B (Fig. 50). Deux amino-acides de ce segment, Arg 10 et Ile 11, interviennent, par des liaisons hydrogène au niveau de la chaîne principale, dans la fixation du groupement pyrophosphate du NAD (Fig. 10), Ile 11 établissant également une interaction hydrophobe avec le noyau nicotinamide.

Le segment 28-32, qui correspond à l'extrémité C-terminale du feuillet β B, est assez fortement conservé. Il renferme notamment le résidu Asp 32 impliqué dans la fixation du ribose de la partie adénylate du cofacteur (Fig. 10).

La région 95-100, qui correspond à une boucle, est assez fortement conservée. Ce fragment renferme notamment les résidus Thr 96 et Phe 99 impliqués dans la fixation de la partie adénine du NAD.

Une forte conservation est observée également dans le segment 111-122 qui correspond au feuillet β E et à une boucle. Aucun amino-acide de cette région n'a été décrit comme intervenant dans la fixation du NAD. La forte conservation de ce long segment suggère qu'il doit jouer néanmoins un rôle essentiel dans le maintien de la structure ou de la fonction de ce domaine de la molécule.

L'extrémité C-terminale du domaine de fixation du coenzyme, c'est-à-dire la région 145-148, présente une assez bonne conservation. Elle se situe à la jonction du feuillet β F et d'une boucle. Elle contient notamment le résidu Ser 148, postulé être impliqué dans la D-spécificité du substrat et participant à la formation du site Pi.

L'hélice α 3 de la région C-terminale appartient au domaine de fixation du coenzyme. L'extrémité N-terminale de cette hélice est extrêmement conservée dans toutes les séquences étudiées. Elle renferme notamment l'Asn 313 et la Tyr 317, impliquées dans la fixation de la partie nicotinamide du cofacteur, respectivement par des interactions électrostatiques et hydrophobes.

Figure 49 : Aligement des structures primaires de GAPDH de différents organismes

	1	10	20	30	40		
a	TKVA	INGFGRIGRLANRR	I	LEVP	---GLEVV	AINDLT-DAKMLAHLFKYD	
b	AV		AA	SRD-S	LT	G-SVEGNFR	
c	AV	G	NVF	AAKN	---DI	V-NTL	
d	M	G	QVF	ILHSRG	---V	AL-NDTL	
e	TI	G	IVF	AAQKRS	---DI	I-L-DYMYML	
f	VR		VM	IASR	---NV	L-PFITNDYAYM	
g	VN	NV	VT	IAISRK	---DINL	PFISTDYAYM	
h	S	IG	VL	AAIDK	---AS	V-PFIVNYMVMF	
i	SKAN	G	VL	AAVKD	---TVQ	V-PFITIDYMYM	
j	S	IG	VL	AA	SCG	---AQ	V-PFIALEYMVM
k	V	GV	VT	AAVLSG	---KVQ	PFI	LNVMVM
l	V	GV	VT	AAFSCD	---KVDI	PFI	LNVMVMQ
m	V	GV	VT	AAFNSG	---KVDI	PFI	LHYMVM
n	GKV	GV	VT	AAFNSG	---KVDI	PFI	LNVMVMQ
o	ADKKI	IG	VA	V	QRN	---DV	LVPFITTEYMTYM
p	GKI	IG	VA	VA	QSE	---DV	LVPFITTDYMTYM
q		VA	VA	QRD	---DV	-VPFISTDYMTYM
rEAKL		NFL	CWHGRGD	ASP	D	I-TG-GVQASL
s	KL		NFL	CWHGRKD	-SP	D	I-TG-GVQASL
t	KL		NFL	CWHGRKD	-SP	D	VV-SG-GVNASL

	50	60	70	80	90					
a	SSQGRF	-NGEIEV-KEGAFVV	---NG-K	---EVKVFAEAD	PEKLPW	GELGIDVVL				
b	AH	TY-P	TVTT-EGNDM	I--D	---KIV	TRANHKKVIM				
c	VH	L-DA	VS-NGN	NL--	---IIK	RNAAIVIV				
d	IYH	-P	VAY-DDQYLY	--D	---AIRAT	VKKEIAAVGI				
e	TH	-D	TV-DHLI	--	---KIR	TRANKDVA				
f	TH	Y-A	VSH-DDKHII	--D	---KIATY	QRANSSNVIAI				
g	TH	-D	VSH-DKDHIIL	--	---KA	NKAAKVAI				
h	VH	-K	TVAA-EGFL	--	-Q---KIT	SRANINASA				
i	TH	Q	-K	TVTY-DGD	FLIQKD	-SSHKI	NSKAAIASVKA			
j	TH	V	-K	VKM-ED	L--D	---KIT	NMKNISKA			
k	TH	H	-K	TVKA-EN	KL	I--H	---AITIQ	RANIKADAAEYV		
l	TH	K	-	TVKA-EN	KL	I--	---PITIQ	RANIKDAAEYV		
m	TH	K	-H	TVKA-ED	KL	I--D	---AITIQ	RANIKDATAYV		
n	TH	K	-H	TVKA-EN	KL	I--	-N---PITIQ	RANIKDAAEYV		
o	H	QWKHN	LK	-DEKTLL	--FE	---PT	GIRN	DI	AA	FV
p	TV	HWKHS	DL	-DSKTLL	--FD	---PT	GIRN	EI	AA	FV
q	VH	QWKHH	LK	-DEKTLL	--FE	---SR	GIRN	EI	AA	FV
r	TL	I	-DADV	KP-GDN	IS	--D	---VI	VSDRN	SN	LI
s	TL	I	-DADV	KPVGID	GIS	--D	---VIQ	VSDRN	VN	LI
t	ML	T	KADV	KI-DNETIS	--D	---HI	VSSR	L	A	II

	200	210	220	230	240	
a	GDFRRARA	AAVSIIPNST	GAAKAI	AQVIPLEL	KGKLDGNAQRV	PVPTGSVTEL
b	- L	SA M T	R V L		ISI	T DV LVDF
c	- L	E	TT	V L L	N M M	T NV VD
d	- L	IN	TT	T L L S	RF M L	TA ISDI
e	- W G	SQN	S	VGK L	N T M F	T NV VD
f	- W G	SGN	S	VGK L	Q T M F	TVDV VD
g	- W GG	T SGN	S	VGK L S	Q T M F	TVDV VD
h	- LW DG	G QN	AA	VGK A	N T M F	T NV VD
i	- LW DG	G GQN	A	VGK	N T M F	T DV VD
j	- W GG	G Q	S	VGK	D T M F	T DV VD
k	- LW DD	QN	A	VGK	N T M F	T NV VD
l	- LW DG	G GN	A	VGK	N T M F	T NA VD
m	- LW DG	G QN	A	VGK	D T M F	T NV VD
n	- LW DG	G LQN	A	VGK	N T M F	T ANV VD
o	- W GG	SFN	S	VGK L	Q N T MSF	TVDV VD
p	- W GG	SFN	S	VGK L	D N T MSF	TVDV VD
q	- W GG	T SFN	S	VGK L	A N T M F	TVDV VD
r	- L	LN V T		VSL L N	N I L	T NV VD
s	- L	LN V T		V LSSQA	R SSMALPL	T NV VD
t	- L	LN V T		VSL L Q	N I L	T NV VD

	250	260	270	280	290	
a	ISVLKKN	-VTVEE	IINAAMKEA	ANE--SL-G	---YTEDEIVS	ADVVGISYGS
b	TF PERD	-T A	SVL A	-DTGDMT	VLG DEPL	R FYSDPHS T
c	VAE E E	- V	L A	-GE K	ILA S EPL	R YN STVS T
d	TAL RE	- A V	L A	-GP K	ILA LZBI	MDPHS I
e	TVR E A	-A Y Q	K V A	-GEMK	VLG DV	T FN EVCT V
f	TVK D E	-T YD	KKVV A	-GK K	VLG AV	S FL D HS I
g	TVN A E	-TSYD	K L K	S-G MK	ILG DV	S FL DAHS I
h	TVR G G	-A YD	K KVE	S-K-GP	K ILG DE V	T FLSDTHS V
i	TVR E P	-ASMDD	KKVV A	-D-GPMK	ILA NV	T F SDPHS I
j	TVR G E	-CSYDD	K T	S-GP Q	FLG DV	S FI DNRS I
k	TCR E P	-AKYDD	KRVV A	-D-GP K	ILG QV	C FN D HS T
l	TCR E P	-AKYDD	KKVV Q	-GP K	ILG QAA	C FNSN HS T
m	TCR E P	-AKYDD	KKVV Q	S-GP K	ILG QV	C FNDSTHS T
n	TCR E P	-AKYDD	KKVV Q	S-GP K	ILG HQV	S FNSDTHS T
o	TVR E A	-A YD	KK I	ES-Q-GK	K ILG DV	T F DNRS I
p	TVR I E	E-ASYDD	K I	ES-GK	K ILGF DV	T F D RS I
q	TVR E G	-ASY D	KK I A	S-GP K	IMG V EDL	T FL D RS I
r	VVQVS	K-TLA	V Q FRD	-AN-E	T ILEVCDVPL	V FRCSDVS T
s	VVQVS	K-TFA	V FR	-DK-E	K ILDVCDEPL	V FRCSDVS T
t	VVNVA	KGI A DV	FRK	-D-GP	K VLAVCDEPL	V FRCSDVS T

	300	310	320	330
a	F D A T L T K I V D V D G S Q L V K T V S H Y D N E M S Y T S Q L V R T L E - Y F A - K I A K .			
b	V S R E A V L E - G - K L - A R V A		W G F - N R M V D T A A Q M - T L .	
c	I L S M V I - G - K M - V		T G - H R V D A A I S G L .	
d	V K A L G - N - M X - X V F A		W G - A N R A D V E L V L R G V .	
e		K A G I A L N - - N F - L	T G - N K L D I A H I - S .	
f		S A G I Q L S - P - K F - L	Y G - T R V D V E H I - A .	
g	V A A G I Q L T - P - T F - L		F G - T R V D V E H V - S .	
h		K A G I S L N - - K F - L I	F G - N R I D I K M Q S D .	
i		G A C I S L N - P - N F - L	Y G - N R V D I G I T R G .	
j		K A G I S L S - K - T F - V	F G - R I D L K H M Q - V D S A .	
k		G A G I A L N - - H F - L	F G - N R V D M V H M S E .	
l		G A G I A L N - - N F - L I	Y G - N R V D M A M S E .	
m		G A G I A L N - - H F - L I	F G - N R V D M V H M S E .	
n		G A G I A L N - - H F - L I	F G - N R V D M A H M S E .	
o		K A G I A L S - - N F - L	W G - T R V D I I H M - S A .	
p		K A G I A L N - - H F - L	W G - N R V D I R H M F - S Q .	
q		K A G I A L S - K - N F - L	W G S R I D I C H M S V A .	
r	I S M V M - - - D D M V I		W G - R V D A D I C N Q W K .	
s	V S M V M - - - D D M V I A		W G - R V D A D I V N Q W K .	
t	I S S M V M - - - D D M V A		W G - R V D A H L V N N W P G S C S T R K .	

a : *C.pasteurianum* (chapitre II)
 b : *Z.mobilis* (Conway et al., 1987)
 c : *B.stearothermophilus* (chapitre IV)
 d : *T.aquaticus* (Hocking et Harris, 1980)
 e : *E.coli* (Branlant et Branlant, 1985)
 f : *S.cerevisiae* (GapA) (Holland et al., 1979)
 g : *Z.rouxii* (Imura et al., 1987)
 h : Drosophile (Gap1) (Tso et al., 1985b)
 i : Nématode (Yarbrough et al., 1987)
 j : Homard (muscle) (Harris et Waters, 1976)
 k : Poulet (muscle) (Dugaiczik et al., 1983)
 l : Rat (foie) (Tso et al., 1985a)
 m : Porc (muscle) (Harris et Perham, 1968)
 n : Homme (foie) (Tso et al., 1985a)
 o : Moutarde cytosol (Martin et Cerff, 1986)
 p : Maïs cytosol (Brinkmann et al., 1987)
 q : Tabac cytosol C (Shih et al., 1986)
 r : Maïs chloroplaste (Brinkmann et al., 1987)
 s : Tabac chloroplaste A (Shih et al., 1986)
 t : Tabac chloroplaste B (Shih et al., 1986)

Z = Glx

B = Asx

X = ?

Les différentes séquences sont alignées sous celle de la GAPDH de *C.pasteurianum*. Seuls les amino-acides différents ont été représentés. L'espace indique la présence d'un amino-acide homologue en cet endroit précis. Le tiret indique une délétion.

Cette hélice contient également le résidu Glu 314 qui participe, par l'intermédiaire d'une molécule d'eau, à la fixation du coenzyme. Nous avons aussi remarqué la forte conservation du segment 129-133, très proche dans l'espace de l'extrémité N-terminale de l'hélice α_3 . Il est possible de suggérer l'existence, entre ces deux régions, d'interactions fondamentales sur le plan structural.

1-2. Les régions fortement divergentes

Une séquence additionnelle de 2 ou 3 résidus est retrouvée spécifiquement entre les positions 23 et 24 des séquences de GAPDH chloroplastiques. Ceci avait déjà été observé par certains auteurs qui avaient alors suggéré que cette insertion pouvait être impliquée dans un élargissement du site de fixation du coenzyme nécessaire pour accueillir le groupement phosphate du NADP, cofacteur spécifique des enzymes chloroplastiques. D'après notre alignement, nous pouvons remarquer qu'une enzyme bactérienne, celle de *Z.mobilis*, renferme le même type d'insertion. De plus, la boucle renfermant cette insertion est localisée dans une région située en surface de la molécule (Fig. 50). Une intervention de cette boucle dans la fixation du coenzyme est peu vraisemblable. D'ailleurs, comme nous l'avons déjà souligné dans la partie "Introduction", une équipe du laboratoire a pu récemment modifier la spécificité de cofacteur de l'enzyme de *B.stearothermophilus* tout en conservant intacte la séquence de cette boucle (Billard, 1988).

La région 123 à 142 présente elle aussi des insertions et délétions, localisées essentiellement dans des boucles.

La très faible conservation de la partie C-terminale de l'hélice α_3 contraste avec la région N-terminale très conservée de cette hélice. Au niveau des variations dans cette région, la sous-unité de type B de l'enzyme chloroplastique de tabac se distingue par la présence d'une longue séquence C-terminale.

1-3. Conclusion

Dans le domaine de fixation du coenzyme, nous avons pu remarquer que des zones très conservées alternaient avec des régions très variables. Les séquences conservées se situent à la fois dans des régions très structurées (hélices α et feuillets β) et dans des boucles.

Un grand nombre des segments conservés renferment des amino-acides intervenant dans la fixation du NAD. Cependant, certains segments, bien que très conservés, ne renferment aucun amino-acide décrit comme intervenant dans la fixation des substrats. Ces segments, tel que le feuillet βE , doivent néanmoins être essentiels au bon fonctionnement de l'enzyme, vraisemblablement en maintenant la structure tridimensionnelle du domaine.

Les régions variables correspondent surtout à des boucles et des feuillets β . Les insertions et les délétions sont presque exclusivement localisées dans des boucles (Fig. 50), situées en surface du tétramère et sur la face opposée à la poche catalytique (Fig. 51). La zone où se situent ces insertions n'intervient pas non plus dans les interactions entre sous-unités. En d'autres termes, les insertions et délétions se sont produites au sein de zones subissant peu de contraintes, aussi bien sur le plan fonctionnel que structural.

D'une manière générale, l'ensemble des segments à séquence peu conservée du domaine de fixation du NAD sont regroupés dans la moitié du domaine opposée au site de fixation du coenzyme, l'ensemble des segments au voisinage de ce site étant eux très conservés (Fig. 50). La chaîne passant successivement d'un côté à l'autre du domaine, ceci explique l'alternance de séquences variables et de séquences conservées le long de la chaîne.

2. Domaine catalytique

Le domaine catalytique de la GAPDH se décompose schématiquement en deux parties, un bloc très conservé entre les résidus 149 et 240 et l'autre présentant une conservation beaucoup plus faible.

2-1. Les régions conservées

Tous les acides aminés importants font l'objet d'une très grande conservation à travers toutes les espèces étudiées. Ainsi, les résidus Thr 150, Arg 195 et Thr 208 intervenant dans la fixation du substrat au site P_i sont tous conservés. Il en est de même pour les résidus Thr 179 et Arg 231 impliqués dans la fixation du substrat au site P_s . Les résidus essentiels à la catalyse, la Cys 149 et l'His 176, sont bien sûr conservés eux aussi. La Cys 149 est localisée dans un segment de 8 amino-acides extrêmement conservés. Dans toutes les séquences établies une seule différence est observée en position 153 dans l'enzyme de *T.aquaticus* dont la séquence, rappelons-le, a été déterminée directement sur la protéine. Ce segment constitue une partie de l'hélice $\alpha 1$. Le résidu His 176 se situe quant à lui dans une région très conservée et qui constitue le feuillet $\beta 1$.

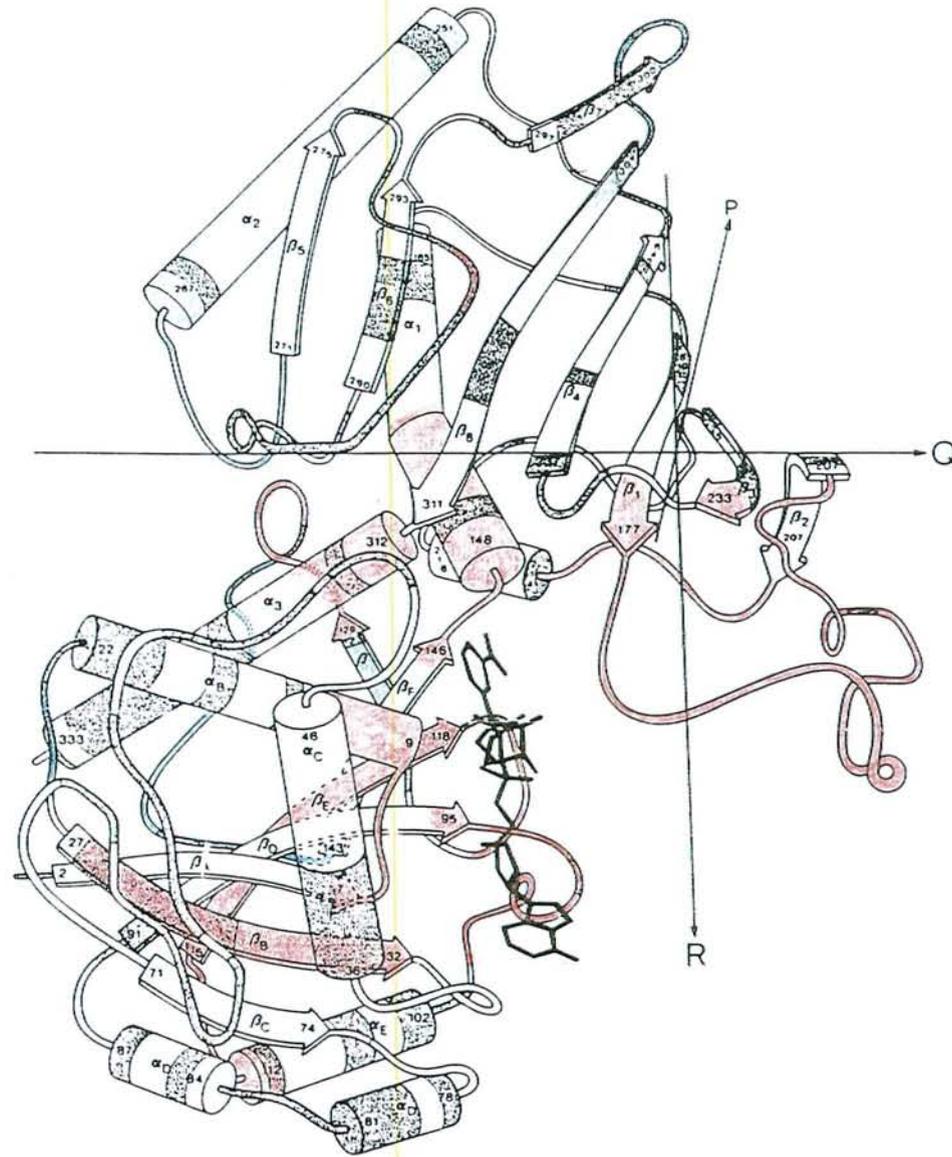


Figure 50 : Localisation des régions conservées et divergentes dans la structure tridimensionnelle de la GAPDH

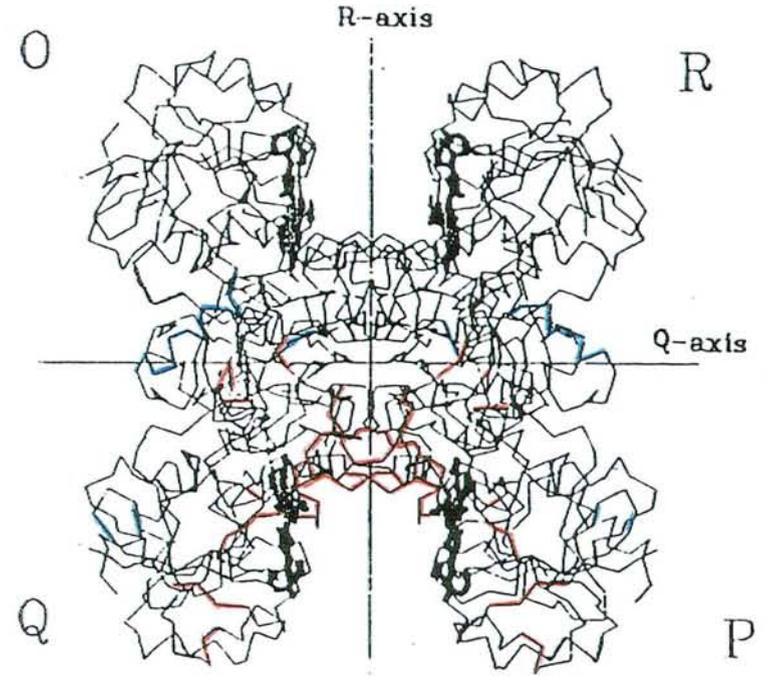


Figure 51 : Les régions conservées dans la structure tétramérique de la GAPDH

Les modèles utilisés sont les structures tridimensionnelles du monomère (Walker et al., 1980) et du tétramère (Skarzynski et Wonacott, 1988) de la GAPDH de *B.stearotherophilus*.

Les régions extrêmement conservées sont indiquées en rouge, les régions très divergentes en bleu. Le degré de conservation a été estimé d'après l'examen des séquences alignées des différentes GAPDH (Fig. 49).

Les autres régions très conservées de ce domaine sont le segment 208-213, qui contribue à la formation du site Pi, et le segment 231-233, qui contribue à la formation du site Ps. Ces régions sont situées respectivement dans une boucle très proche de la "S-loop" dans l'espace et à l'extrémité C-terminale du feuillet β_3 .

Il faut noter également la présence d'un court segment très conservé autour de la Ser 280. Certains résidus de ce segment pourraient être impliqués dans des interactions stabilisatrices de la structure tétramérique de la GAPDH.

2-2. Les régions variables

Les zones variables se situent, comme nous l'avons déjà indiqué, dans la deuxième moitié du domaine catalytique. Elles concernent à la fois des boucles, des feuillets β et des hélices α , en particulier le segment 251-267 présentant une faible conservation et qui constitue l'hélice α_2 . Notons que cette hélice est localisée très loin du site actif à la périphérie de la molécule.

Les nombreuses insertions et délétions constatées au niveau des deux régions 263-270 et 301-305 sont situées essentiellement dans des boucles localisées elles aussi en surface de la protéine. Le segment 263-270 présente de nombreuses délétions dans l'enzyme de *C.pasteurianum*, délétions qui ne sont retrouvées dans aucune autre espèce. Le segment 301-305 de cette même enzyme présente lui au contraire trois insertions relevées aux positions 301A, 302A et 304A, des insertions et délétions étant également retrouvées dans cette région dans les enzymes chloroplastiques.

2-3. Conclusion

Cette étude montre que, comme dans le cas du domaine de fixation du coenzyme, les régions très conservées dans le domaine catalytique sont essentiellement localisées dans les parties centrales du tétramère constituant la poche catalytique (Fig. 50 et 51). Ces régions interviennent dans des fonctions catalytiques ou structurales importantes.

Les régions présentant le plus de divergences, d'insertions et délétions, sont elles disposées à la surface de la molécule, loin du site actif.

III - LA FILIATION DES ESPECES D'APRES L'ETUDE DE LA GAPDH

1. Comparaisons des GAPDH bactériennes

Les enzymes bactériennes présentent entre elles des pourcentages d'homologie compris entre 49 et 62 %, valeurs déterminées en comparant les structures primaires des protéines dans leur ensemble (Tab. III). Les pourcentages les plus élevés sont observés avec l'enzyme de *B.stearothermophilus* qui présente curieusement une forte homologie avec toutes les autres enzymes bactériennes.

1-1. Comparaison entre *E.coli* et les quatre bactéries Gram+

Comme nous l'avons déjà mentionné dans la partie "Introduction", les études précédentes avaient montré qu'*E.coli* était plus proche des enzymes cytosoliques d'eucaryotes que de celles de *B.stearothermophilus* (Branlant et Branlant, 1985) ou des chloroplastes (Martin et Cerff, 1986).

L'alignement que nous avons réalisé (Fig. 49) et les pourcentages d'homologie que nous avons calculés (Tab. III à VI) montrent qu'elle est également plus proche des enzymes eucaryotiques que des enzymes de bactéries Gram+ mésophiles. La GAPDH d'*E.coli* se distingue donc de toutes les GAPDH bactériennes étudiées. Il serait nécessaire à présent de connaître la séquence de la GAPDH d'une autre bactérie Gram- afin de déterminer si cette situation est spécifique à *E.coli* ou si elle est caractéristique de l'ensemble des bactéries Gram-.

Comme nous l'avons vu précédemment, la forte homologie de l'enzyme d'*E.coli* par rapport aux enzymes cytosoliques d'eucaryotes se situe essentiellement au niveau du domaine catalytique (72 à 82 % d'homologie). Au niveau du domaine de fixation du coenzyme, les pourcentages d'homologie par rapport aux enzymes eucaryotes ne sont pas significativement plus élevés que ceux obtenus par rapport aux autres enzymes bactériennes (de 47 à 60 % avec les enzymes cytosoliques, de 46 à 58 % avec les enzymes des bactéries Gram+).

1-2. Comparaison entre les GAPDH de bactéries Gram+ mésophiles et thermophiles

L'alignement de la figure 49 et les résultats consignés dans les tableaux III à VI montrent clairement que les enzymes des deux bactéries thermophiles présentent entre elles un degré de conservation plus fort que les enzymes des bactéries mésophiles.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	
B A C T E R I E S	a. <i>C.pasteur</i>	**																			
	b. <i>Z.mobilis</i>	50	**																		
	c. <i>B.stearo</i>	62	59	**																	
	d. <i>T.aquatic</i>	52	53	60	**																
	e. <i>E.coli</i>	53	51	58	49	**															
L E V	f. <i>S.cerevis</i>	45	47	50	45	70	**														
	g. <i>Z.rouxii</i>	47	46	50	46	66	83	**													
	h. Drosophile	46	47	53	48	66	65	61	**												
	i. Nématode	46	46	54	49	65	65	66	76	**											
A N I M A L S	j. Homard	48	47	52	48	64	66	67	76	70	**										
	k. Poulet	45	46	52	48	68	66	66	77	73	73	**									
	l. Rat	46	45	49	47	65	65	66	75	72	71	92	**								
	m. Porc	45	47	51	47	68	65	66	77	74	72	92	92	**							
	n. Homme	45	46	51	46	67	63	64	75	72	70	91	93	93	**						
	o. Mout.cyt.	46	47	50	45	64	68	68	66	69	68	67	66	67	**						
V E G E T A L S	p. Maïs cyt.	45	46	49	46	65	68	67	69	69	67	71	67	70	70	86	**				
	q. Tabac cyt.	45	45	50	46	63	67	67	67	69	67	67	65	66	66	84	80	**			
	r. Maïs chl.	45	47	59	52	48	51	48	50	48	44	47	48	48	47	47	44	46	**		
	s. Tab.chl.A	44	46	56	48	46	50	48	50	47	42	46	45	48	46	45	43	44	91	**	
	t. Tab.chl.B	45	48	57	52	46	51	48	49	48	43	45	46	45	46	45	44	46	78	78	**

Tableau III : Pourcentages d'homologie totale entre les GAPDH

(Les valeurs indiquées ont été calculées à partir de l'alignement des séquences représenté par la figure 49.)

Lorsque les séquences sont comparées dans leur ensemble, les pourcentages d'homologie sont de 60 % pour les enzymes thermophiles et de 50 % pour les enzymes mésophiles (Tab. III).

Nous avons tout de même constaté que les séquences des enzymes de *C.pasteurianum* et de *Z.mobilis* possèdent, en une dizaine de positions, des acides aminés identiques, les mêmes positions étant occupées par des amino-acides différents dans les autres séquences. Il en est ainsi des résidus Ala 4, Ala 15, Ile 19, Gly 24, Leu 25, Leu 87, Cys 95, Leu 136 et Val 245.

La conservation des enzymes de thermophiles est très élevée dans la région de la "S-loop" (Fig. 52), le pourcentage d'homologie étant de 87 % alors qu'il est seulement de 58 % entre les "S-loop" des deux enzymes de mésophiles (Tab. VI). Néanmoins, les "S-loop" des GAPDH de thermophiles ne présentent qu'un seul amino-acide qui leur soit réellement caractéristique, le résidu Leu 187. Ceci semble indiquer que la "S-loop" n'est peut-être pas aussi directement impliquée dans la thermostabilité des GAPDH que cela avait été suggéré. D'autre part, tenant compte du fait que les résidus Asn 180 et Leu 185 étaient présents chez *B.stearotherophilus* et *T.aquaticus* mais pas dans les enzymes eucaryotiques, Skarzynski et Wonacott (1988) avaient suggéré que ces résidus pourraient être impliqués dans des interactions entre les différentes sous-unités de la GAPDH et ainsi contribuer à la stabilité thermique des enzymes de thermophiles. Cette hypothèse semble à présent devoir être écartée, les résidus en question étant retrouvés également dans les séquences de deux bactéries mésophiles, *C.pasteurianum* et *Z.mobilis*.

1-3. Conclusion

Comme le montre le tableau III, les GAPDH bactériennes présentent entre elles une conservation nettement plus faible (de 49 à 62 %) que les enzymes eucaryotiques (de 63 à 93 %) ou les enzymes chloroplastiques (de 78 à 91 %).

Par son homologie très marquée avec les enzymes cytosoliques des eucaryotes, l'enzyme d'*E.coli* reste un cas particulier. Les deux autres enzymes mésophiles présentent avec l'enzyme d'*E.coli* et entre elles de nombreuses divergences, surtout localisées dans les parties du tétramère éloignées du site catalytique.

		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
B A C T E R I E S	a. <i>C.pasteur</i>	**	76	69	66	52	52	52	59	66	69	62	59	62	62	69	69	69	55	55	55
	b. <i>Z.mobilis</i>	51	**	69	69	52	48	52	66	69	69	66	62	66	66	66	66	66	59	59	55
	c. <i>B.stearo</i>	60	49	**	72	52	55	59	69	69	66	72	62	72	72	66	66	66	59	59	59
	d. <i>T.aquatic</i>	51	48	56	**	48	62	62	72	72	66	69	66	69	69	76	76	76	62	62	69
	e. <i>E.coli</i>	50	46	58	50	**	62	59	48	48	55	48	45	48	48	48	48	48	59	59	55
L E V	f. <i>S.cerevis</i>	43	45	50	45	59	**	83	59	55	59	52	48	52	52	62	62	62	62	62	62
	g. <i>Z.rouxii</i>	44	45	50	46	60	76	**	55	52	55	52	48	52	52	59	59	59	59	59	59
A N I M A U X	h. Drosophile	44	43	52	49	52	56	55	**	86	86	83	79	83	79	83	83	83	52	52	55
	i. Nématode	45	44	49	47	54	58	57	70	**	83	83	79	83	86	83	83	83	55	55	59
	j. Homard	47	47	49	47	53	58	58	78	62	**	83	79	83	79	83	83	83	55	55	52
	k. Poulet	45	46	46	48	54	53	53	73	60	72	**	97	100	97	79	79	79	55	55	55
	l. Rat	46	47	50	48	54	56	58	71	60	70	89	**	97	93	76	76	76	52	52	52
	m. Porc	43	46	48	49	56	55	55	70	65	68	90	93	**	97	79	79	79	55	55	55
	n. Homme	42	42	47	48	56	52	52	68	60	65	90	92	91	**	76	76	76	59	59	59
	o. Mout.cyt.	42	40	46	43	48	56	54	56	58	62	56	56	57	56	**	100	100	55	55	55
V E G E T A U X	p. Maïs cyt.	41	40	44	41	48	55	53	60	57	60	60	58	60	59	83	**	100	55	55	55
	q. Tabac cyt.	43	41	45	42	47	53	55	57	57	57	58	56	58	55	82	78	**	55	55	55
	r. Maïs chl.	44	44	46	45	45	49	45	50	45	47	43	46	46	45	43	41	42	**	100	86
	s. Tab.chl.A	44	43	47	42	44	49	47	50	45	46	43	47	48	46	41	40	41	92	**	86
t. Tab.chl.B	45	43	47	48	41	48	45	48	45	43	40	43	43	42	40	41	41	75	74	**	

Tableau IV : Pourcentages d'homologie entre les domaines de fixation du coenzyme des GAPDH

Les résultats indiqués en-dessous de la diagonale correspondent aux comparaisons effectuées sur la séquence primaire du domaine entier. Les résultats au-dessus de la diagonale correspondent à la région comprise entre les résidus Glu 94 et Ala 122.

(Les valeurs indiquées ont été calculées à partir de l'alignement des séquences représenté par la figure 49.)

		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
B	a. <i>C.pasteur</i>	**	65	71	63	63	61	55	57	59	65	63	61	61	57	67	67	67	77	65	77
A																					
C	b. <i>Z.mobilis</i>	49	**	75	61	65	65	59	55	61	65	61	57	59	57	63	63	59	71	61	73
T																					
E	c. <i>B.stearo</i>	64	69	**	73	77	73	67	69	67	73	73	69	71	67	67	67	67	84	71	84
R																					
I	d. <i>T.aquatic</i>	53	58	64	**	59	59	57	53	51	53	55	63	53	53	55	55	55	69	59	73
E																					
S	e. <i>E.coli</i>	56	57	59	48	**	92	86	82	84	86	88	82	80	80	84	84	84	71	57	71
L																					
E	f. <i>S.cerevis</i>	47	49	57	45	81	**	94	75	80	84	80	78	78	78	88	88	88	69	55	69
V																					
	g. <i>Z.rouxii</i>	49	46	50	46	72	90	**	77	80	84	77	78	78	77	88	88	88	65	51	65
	h. <i>Drosophile</i>	48	51	54	47	80	74	73	**	92	84	92	92	94	92	77	77	77	59	49	63
	i. Nématode	47	48	59	51	76	72	75	81	**	88	92	92	94	92	77	77	77	59	47	61
A																					
N	j. Homard	48	47	55	48	82	74	76	74	78	**	84	84	90	84	82	82	82	67	55	67
I																					
M	k. Poulet	45	46	54	48	82	76	74	83	86	73	**	92	94	92	78	78	78	71	59	71
A																					
U	l. Rat	44	43	48	47	76	72	74	79	84	70	94	**	94	92	77	77	77	67	55	67
X																					
	m. Porc	47	48	54	46	78	71	73	80	79	72	92	92	**	94	77	77	77	69	57	69
	n. Homme	47	50	55	44	78	74	76	83	84	75	92	93	95	**	78	78	78	67	55	67
	o. Mout.cyt.	50	54	54	47	80	80	82	76	80	74	77	76	77	77	**	98	94	67	53	69
V																					
E	p. Maïs cyt.	49	52	54	51	82	81	80	78	81	74	81	76	79	80	90	**	94	67	53	67
G																					
E	q. Tabac cyt.	47	49	55	50	79	81	78	76	80	77	77	73	75	76	88	83	**	67	55	69
T																					
A	r. Maïs chl.	46	50	71	59	51	53	51	50	52	41	51	48	50	49	51	47	51	**	77	98
U																					
X	s. Tab.chl.A	44	49	65	54	48	51	49	49	49	38	49	44	47	46	48	45	48	90	**	77
	t. Tab.chl.B	45	53	67	56	51	53	52	50	51	42	52	47	49	48	51	48	51	81	81	**

Tableau V : Pourcentages d'homologie entre les domaines catalytiques des GAPDH

Les résultats indiqués en-dessous de la diagonale correspondent aux comparaisons effectuées sur la séquence primaire du domaine entier. Les résultats au-dessus de la diagonale correspondent à la région comprise entre les résidus Asp 192 et Leu 242.

(Les valeurs indiquées ont été calculées à partir de l'alignement des séquences représenté par la figure 49.)

2. Comparaisons entre les GAPDH bactériennes et les GAPDH chloroplastiques

Martin et Cerff (1986) avaient constaté que les enzymes de *B.stearotherophilus* et *T.aquaticus* présentaient une homologie plus forte avec les GAPDH chloroplastiques qu'avec les enzymes cytosoliques des mêmes organismes végétaux. Ceci les avait conduits, ainsi que Shih et al. (1986), à suggérer que les chloroplastes provenaient d'un ancêtre procaryotique symbiotique dont les gènes auraient été transférés vers le noyau de la cellule-hôte végétale au cours de l'évolution.

Le tableau III montre bien que les enzymes chloroplastiques présentent de fortes homologies avec celles des bactéries, et en particulier celles des bactéries thermophiles. L'homologie la plus élevée est observée avec l'enzyme de *B.stearotherophilus*, malgré la correction de 20 amino-acides que nous avons apportée en déterminant la séquence nucléotidique du gène de cette enzyme. Seule l'enzyme d'*E.coli* présente une homologie plus importante avec les enzymes végétales du cytosol qu'avec celles du chloroplaste.

Si nous examinons les pourcentages d'homologie entre les enzymes de thermophiles et les enzymes de chloroplastes au niveau de chaque domaine, nous observons là encore que le degré d'homologie est beaucoup plus élevé au niveau du domaine catalytique (65 à 71 %) (Tab. V). Au niveau du domaine de fixation du coenzyme, les degrés d'homologie des enzymes chloroplastiques et des enzymes bactériennes thermophiles sont sensiblement identiques à ceux observés avec les enzymes cytosoliques d'eucaryotes. Plus précisément, au sein du domaine catalytique, c'est dans la "S-loop" que l'homologie entre les enzymes bactériennes et chloroplastiques est la plus marquante (66 % avec les enzyme mésophiles, 74 % avec les enzymes thermophiles). Notons que ce pourcentage chute à 47 % pour l'enzyme d'*E.coli*, pourcentage similaire à celui observé avec les enzymes cytosoliques (Tab. VI).

Les pourcentages d'homologie observés s'expliquent par l'existence d'amino-acides présents uniquement dans les "S-loop" des enzymes des bactéries Gram⁺ et des enzymes chloroplastiques : Tyr 178, Asp 181, Arg 183, Leu 185, Phe 193 et Arg 195, ces amino-acides n'étant pas présents dans la "S-loop" de l'enzyme d'*E.coli*.

D'autres positions hors de la "S-loop" sont également occupées par des amino-acides identiques dans les enzymes chloroplastiques et bactériennes, alors qu'elles sont occupées par d'autres acides aminés dans les autres enzymes (Asp 181, Arg 183, Leu 185, His 190, Leu 193, Leu 216, Ala 252, Asn 256, Leu 296 et Thr 297). De même, le résidu en position 35 a fait l'objet d'une délétion retrouvée dans les enzymes chloroplastiques comme dans les enzymes bactériennes.

	180				190				200																					
a	Y	T	N	D	Q	N	T	L	D	G	P	-	H	R	K	G	D	F	R	R	A	R	A	A	A	V				
b	F				R	I			Q	I	-	-	S	-	L										S	A				
c					R	I			L	-	-	-	-	L													E			
d					N	R	L		L	-	-	-	-	L													I			
e	T	A	T	K	V				S	-	-	-	W	G													S	Q		
f	L	A	T	K	V				S	-	-	-	W	G														S	G	
g	M	A	T	K	V				S	-	-	-	W	G	G				T								S	G		
h	T	A	T	K	V				S	G	-	-	L	W	D	G			G									Q		
i	V	A	T	K	V				S	G	-	-	L	W	D	G			G									G	Q	
j	V	A	T	K	V				S	A	-	-	-	W	G	G			G									Q		
k	I	A	T	K	V				S	G	-	-	L	W	D	D												Q		
l	I	A	T	K	V				S	G	-	-	L	W	D	G			G									G		
m	I	A	T	K	V				S	G	-	-	L	W	D	G			G									Q		
n	I	A	T	K	V				S	G	-	-	L	W	D	G			G									L	Q	
o	I	A	T	K	V				S	M	-	-	-	W	G	G												S	F	
p	I	A	T	K	V				S	A	-	-	-	W	G	G													S	F
q	L	A	T	K	V				S	M	-	-	-	W	G	G												T	S	F
r		G			R	L			A	-	S	-	-	-	L														L	
s		G			R	L			A	-	S	-	-	-	L														L	
t		G			R	L			A	-	S	-	-	-	L														L	

a : *C.pasteurianum*
b : *Z.mobilis*
c : *B.stearothermophilus*
d : *T.aquaticus*
e : *E.coli*
f : *S.cerevisiae* (GapA)
g : *Z.rouxii*
h : *Drosophile* (Gap1)
i : Nématode
j : Homard (muscle)
k : Poulet (muscle)
l : Rat (foie)
m : Porc (muscle)
n : Homme (foie)
o : Moutarde cytosol
p : Maïs cytosol
q : Tabac cytosol C
r : Maïs chloroplaste
s : Tabac chloroplaste A
t : Tabac chloroplaste B

Figure 52 : Alignement des séquences de "S-loop" des GAPDH

		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
B	a. <i>C.pasteur</i>	**																			
A																					
C	b. <i>Z.mobilis</i>	58	**																		
T																					
E	c. <i>B.stearo</i>	75	74	**																	
R																					
I	d. <i>T.aquatic</i>	74	65	87	**																
E																					
S	e. <i>E.coli</i>	57	51	51	51	**															
L	f. <i>S.cerevis</i>	57	47	51	51	92	**														
E																					
V	g. <i>Z.rouxii</i>	49	38	43	43	85	88	**													
	h. Drosophile	45	26	38	38	75	67	71	**												
	i. Nématode	41	26	34	34	71	67	79	92	**											
A																					
N	j. Homard	49	30	43	43	83	75	88	83	83	**										
I																					
M	k. Poulet	41	30	43	43	79	71	75	88	88	75	**									
A																					
U	l. Rat	49	26	38	38	71	71	83	92	88	79	88	**								
X																					
	m. Porc	45	26	38	38	75	67	79	96	92	83	92	96	**							
	n. Homme	45	26	34	34	75	67	79	92	92	79	88	92	96	**						
	o. Mout.cyt.	49	38	43	38	88	83	82	71	71	79	79	75	71	75	**					
V	p. Maïs cyt.	49	38	43	38	88	83	83	71	71	83	79	75	71	75	92	**				
E																					
G	q. Tabac cyt.	45	34	38	34	83	83	79	67	67	75	75	67	63	67	88	88	**			
E																					
T	r. Maïs chl.	66	65	74	74	47	47	38	34	26	34	38	30	30	26	34	34	30	**		
A																					
U	s. Tab.chl.A	66	65	74	74	47	47	38	34	26	34	38	30	30	26	34	34	30	100	**	
X																					
	t. Tab.chl.B	66	65	74	74	47	47	38	34	26	34	38	30	30	26	34	34	30	100	100	**

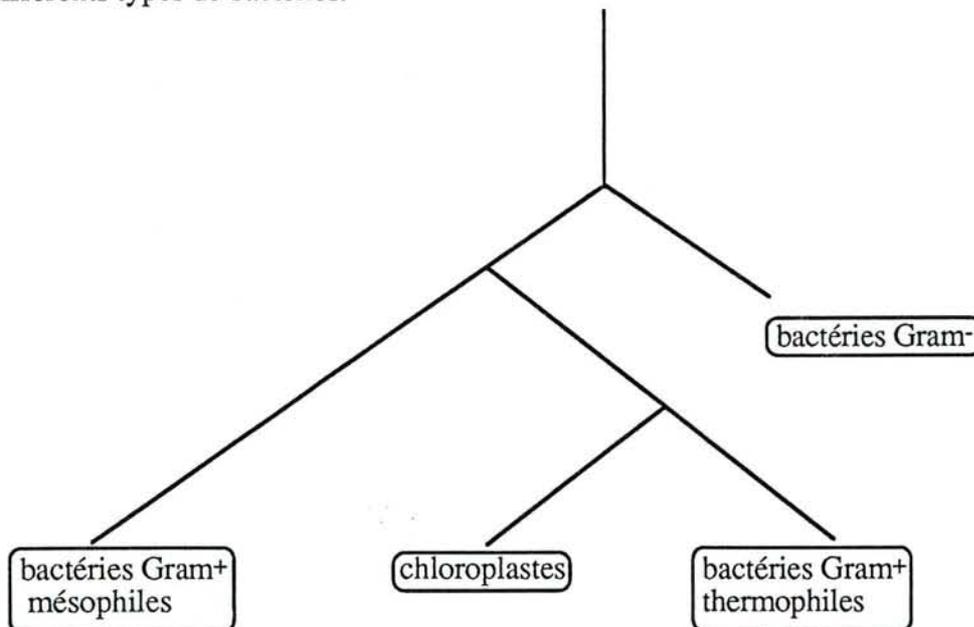
Tableau VI : Pourcentages d'homologie entre les "S-loop" des GAPDH

(Les valeurs indiquées ont été calculées à partir de l'alignement des séquences représenté par la figure 52.)

Comme l'avaient avancé Shih et al. (1986), il semble donc que les chloroplastes aient évolué à partir d'un ancêtre commun avec les bactéries thermophiles. Cette hypothèse repose essentiellement sur les observations actuelles réalisées à partir du domaine catalytique de la GAPDH des différentes espèces, puisque nous avons vu qu'il est impossible de tirer des enseignements significatifs sur le plan de l'évolution à partir du domaine de fixation du coenzyme.

Les résultats obtenus dans le cas du domaine de fixation du NAD s'expliquent sans doute par le fait qu'au niveau de ce domaine, la pression pour la conservation des acides aminés est moins importante. Comme nous l'avons vu, la chaîne passant alternativement de la zone conservée à la zone variable, ce domaine ne présente pas une longue séquence conservée comme la "S-loop" du domaine catalytique où des comparaisons peuvent être réalisées.

Il est possible de proposer un arbre phylogénétique établissant la filiation entre les chloroplastes et les différents types de bactéries.



3. Comparaisons entre les GAPDH bactériennes et le GAPDH cytosoliques d'eucaryotes

Le tableau III montre que, dans l'ensemble, les enzymes bactériennes présentent une homologie d'environ 50 % avec les enzymes cytosoliques animales ou végétales.

Comme nous l'avons déjà souligné, l'enzyme d'*E.coli* se distingue des autres enzymes bactériennes en présentant avec les enzymes eucaryotiques un pourcentage d'homologie d'environ 67 %, comparable à ceux observés en comparant entre elles les enzymes des eucaryotes inférieurs (Tab. III).

Un certain nombre d'acides aminés trouvés uniquement dans les enzymes cytosoliques ont également été retrouvés dans l'enzyme d'*E.coli*.

Au niveau de la région de la "S-loop", l'enzyme d'*E.coli* présente, avec les enzymes cytosoliques, des pourcentages d'homologie variant entre 71 et 92 %, tandis que les pourcentages d'homologie des autres enzymes bactériennes sont compris entre 26 et 57 %.

De toute évidence, l'enzyme d'*E.coli* présente des caractéristiques communes avec les enzymes cytosoliques avec lesquelles elles possèdent une grande analogie structurale.

De nombreuses hypothèses ont tenté d'expliquer la parenté phylogénétique apparemment proche d'*E.coli* avec les eucaryotes. L'une d'entre elles, avancée par Martin et Cerff (1986), suggérait que des gènes eucaryotiques aient pu subir un transfert vers *E.coli*, par l'intermédiaire de la transcription inverse, procédé supposé conduire à des pseudogènes chez les eucaryotes (Sharp, 1983 ; Lewin, 1983). L'analyse de la séquence de la GAPDH d'une autre bactérie Gram⁻ est nécessaire pour savoir si cette hypothèse est plausible.

IV - CONCLUSIONS

Les études comparatives que nous avons exposées sont les premières à prendre en compte cinq séquences de GAPDH bactériennes, comprenant notamment les séquences des enzymes de *C.pasteurianum* et *Z.mobilis*, ainsi que la séquence corrigée de *B.stearothermophilus*.

Nous avons ainsi constaté que la forte homologie présente entre l'enzyme d'*E.coli* et les enzymes cytosoliques des eucaryotes n'était pas due au caractère mésophile de la bactérie, cette forte homologie n'étant pas retrouvée en comparant les enzymes des autres bactéries mésophiles, *C.pasteurianum* et *Z.mobilis*, aux enzymes cytosoliques.

Nous avons observé une homologie élevée entre les enzymes chloroplastiques et les enzymes des bactéries Gram⁺, ce qui confirme l'origine bactérienne des chloroplastes. Parmi l'ensemble des bactéries Gram⁺, nos études ont montré que ce sont les enzymes de thermophiles qui étaient les plus semblables aux enzymes des chloroplastes, suggérant une évolution possible à partir d'un ancêtre commun.

Bien que nous l'ayons peu évoqué dans ce chapitre, il nous faut rappeler que lors de l'évolution des chloroplastes à partir d'un ancêtre procaryotique, l'enzyme chloroplastique a nécessairement dû subir des contraintes qui ont conduit l'enzyme à assumer sa fonction dans la photosynthèse en s'accommodant des conditions propres aux chloroplastes, comme la présence importante de NADPH qui a certainement influé sur la spécificité de coenzyme des GAPDH chloroplastiques. Cependant, il est probable que les modifications induites dans la structure primaire de ces enzymes soient peu nombreuses, ce qui doit se traduire par des variations minimales des pourcentages d'homologie obtenus en comparant les séquences entières.

Afin de vérifier si nos conclusions relatives aux comparaisons entre les enzymes chloroplastiques et les enzymes bactériennes ne sont pas biaisées par le fait qu'elles font intervenir des enzymes anaérobies strictes, il sera nécessaire de déterminer la structure primaire de la GAPDH d'une bactérie mésophile Gram⁺ aérobie. Nous devrions très prochainement apporter de nouveaux éléments de réflexion sur ce sujet, puisque nous venons de cloner le gène de la GAPDH de *Streptococcus lactis* subsp. *diacetylactis*.

BILAN ET PERSPECTIVES

Comme nous l'avons déjà souligné à plusieurs reprises, la GAPDH est une enzyme constituant un bon modèle pour l'étude de l'évolution des espèces.

Nous avons cloné et séquencé les gènes de la GAPDH des bactéries Gram⁺ *C.pasteurianum* et *B.stearothermophilus*. Nous avons ainsi pu étudier l'évolution phylogénétique de cette enzyme à partir de la connaissance de la structure primaire des cinq GAPDH bactériennes connues à ce jour. Nous en avons tiré les conclusions suivantes.

1 - Le domaine de fixation du NAD et le domaine catalytique n'évoluent pas de la même façon. Le domaine catalytique est plus informatif au niveau de l'étude de la filiation des espèces.

2 - Les enzymes bactériennes présentent entre elles une divergence relativement importante, comparativement à ce qui est observé pour les enzymes cytosoliques d'eucaryotes ou pour les enzymes chloroplastiques.

3 - L'enzyme d'*E.coli* représente un cas particulier, puisqu'elle possède plus d'homologies avec les enzymes cytosoliques des eucaryotes qu'avec les autres enzymes bactériennes. Cependant, il est nécessaire de déterminer la séquence d'une autre enzyme de bactérie Gram⁻ afin de savoir si cette particularité phylogénétique est spécifique d'*E.coli* ou si elle est, au contraire, générale à l'ensemble des bactéries Gram⁻.

4 - Les enzymes chloroplastiques présentent une forte homologie avec les enzymes bactériennes, confirmant ainsi la théorie endosymbiotique de l'origine des chloroplastes. Nous avons constaté que ce sont les enzymes des bactéries thermophiles qui se rapprochent le plus des enzymes chloroplastiques, ce qui permet de suggérer l'existence d'un ancêtre commun aux enzymes bactériennes et aux enzymes des chloroplastes.

Des comparaisons plus précises nous ont conduits à observer que le degré de conservation des différentes régions de la GAPDH dépend très fortement de leur localisation dans la structure tridimensionnelle de l'enzyme. C'est ainsi que nous avons observé une très grande conservation de la structure primaire de l'enzyme dans la région constituant la poche catalytique. Les régions divergentes, en particulier celles où se sont produites un grand nombre d'insertions et de délétions, appartiennent au contraire à des zones situées à la surface de la molécule, éloignées du site catalytique. De plus, les insertions et délétions se produisent en majorité dans des boucles.

Il sera important de tenir compte de ces observations dans de futures études visant à prédire la structure d'enzymes dont la structure cristallographique n'a pas été déterminée, mais dont la

structure primaire de plusieurs espèces est connue. Entre les différents modèles possibles, celui dans lequel les variations en tailles se situeront dans des boucles sera le plus favorable.

Notre étude a apporté des informations intéressantes dans un autre domaine, celui de l'expression de gènes hétérologues chez *E.coli*.

1 - Nous avons montré qu'il était possible, en utilisant chez *E.coli* des promoteurs de bactéries Gram⁺, et même de bactéries Gram⁺ thermophiles, d'obtenir une expression efficace chez cet hôte.

2 - Nous avons constaté que l'usage des codons ne produisait pas d'effet dramatique sur l'expression d'un gène. Cet élément constitue donc vraisemblablement un paramètre qui, bien que non négligeable, ne doit pas être déterminant dans la régulation du taux d'expression. Néanmoins, l'usage des codons intervient sans doute également dans la fidélité de la traduction (ce point n'a pas été examiné dans notre étude).

Enfin, notre travail a permis d'approfondir les connaissances sur la structure des gènes de *C.pasteurianum* et *B.stearothermophilus* dont peu de gènes avaient été étudiés jusqu'à présent.

C'est ainsi que nous avons pu observer que ces deux bactéries présentaient en commun une structure en tandem associant les gènes de la GAPDH et de la PGK. Il pourrait s'agir d'un opéron. La même organisation a été retrouvée chez *Z.mobilis*, une autre bactérie Gram⁺. Par contre, cette structure en tandem n'existe pas chez *E.coli*.

Nous avons suggéré que l'organisation en tandem des gènes de la GAPDH et de la PGK pourrait être caractéristique des bactéries Gram⁺. Cependant, en tenant compte de la particularité de l'enzyme d'*E.coli* observée lors de l'étude phylogénétique, il est possible de penser que le gène de la GAPDH ait pu être introduit dans le génome d'*E.coli* à la suite d'un transfert génétique à partir d'un hôte eucaryotique. Aussi, afin de savoir si cette structure en tandem n'existe pas chez les bactéries Gram⁻, il est nécessaire d'étudier ces gènes chez une autre bactérie Gram⁻.

Il serait intéressant de cloner les gènes de la PGK, à partir des banques génomiques de *C.pasteurianum* et de *B.stearothermophilus*, en utilisant comme sondes les fragments de gènes clonés dans les plasmides utilisés pour ce travail. Ceci permettrait de compléter la séquence de ces gènes et de construire une structure en tandem complète afin de pouvoir débiter des expériences de transcription et de transcription-traduction *in vitro*.

Celles-ci devraient permettre de localiser expérimentalement le site d'initiation de la transcription situé en amont du gène de la GAPDH et donc de l'opéron. Ceci sera particulièrement intéressant dans le cas de *B.stearothermophilus*, où nous suggérons l'existence d'un promoteur très éloigné de la région codante. Il sera également possible de mettre en évidence d'éventuels sites internes de réinitiation. Une étude de la régulation de l'expression des gènes *in vivo* chez *B.stearothermophilus* ou chez *C.pasteurianum* en utilisant les fragments clonés comme sonde pourra être envisagée.

Les expériences de transcription et transcription-traduction *in vitro* permettront également d'étudier plus en détail la phase ouverte de lecture G3 du plasmide pCGap1 de *C.pasteurianum* pour laquelle beaucoup de questions restent posées.

Enfin, des informations supplémentaires concernant l'organisation en opéron, mais également la structure et l'expression des gènes de bactéries Gram⁺ et l'évolution phylogénétique des GAPDH bactériennes, pourront être apportées avec l'étude de la GAPDH de *S.lactis* subsp. *diacetylactis*, bactérie Gram⁺ mésophile aérobie, dont nous venons de cloner le gène.

REFERENCES
BIBLIOGRAPHIQUES

ALGAR E.M. et SCOPES R.K. (1985)

Studies on cell-free metabolism : ethanol production by extracts of *Zymomonas mobilis*.
J. Biotechnol., **2** , pp 275-287.

ARCARI P., MARTINELLI R. et SALVATORE F. (1984)

The complete sequence of a full length cDNA for human liver glyceraldehyde-3-phosphate dehydrogenase : evidence for multiple mRNA species.
Nucleic Acids Res., **12**, pp 9176-9189.

ASHMARINA L.I., MURONETZ V.I. et NAGRADOVA N.K. (1985)

Yeast glyceraldehyde-3-phosphate dehydrogenase. Evidence that subunit cooperativity in catalysis can be controlled by the formation of a complex with phosphoglycerate kinase.
Eur. J. Biochem., **149**, pp 67-72.

ASSOBHEI O. (1988)

Etudes sur la glyceraldéhyde-3-phosphate déshydrogénase de *Clostridium pasteurianum* : isolement et structure d'un fragment de 3,7 kb contenant son gène chez *Escherichia coli* et étude des caractères de l'enzyme.
 Thèse d'Université. Nancy 1.

ATKINSON T., BANKS G.T., BRUTON C.J., CONNER M.J., JAKES R., KAMALAGHARAN T., WHITAKER A.R. et WINTER G.P. (1979)

Large scale isolation of enzymes from *Bacillus stearothermophilus*.
J. Appl. Biochem., **1**, pp 247-258.

BACHMANN B.J. (1983)

Linkage map of *Escherichia coli* K-12.
Microbiol. Revs., **47**, pp 180-230.

BAMBERGER E.S., EHRLICH B.A. et GIBBS M. (1975)

The glyceraldehyde 3-phosphate and glycerate 3-phosphate shuttle and carbon dioxide assimilation in intact spinach chloroplasts.
Plant Physiol., **55**, pp 1023-1030.

BARKER D.G. (1982)

Cloning and amplified expression of the tyrosyl-tRNA synthetase genes of *Bacillus stearothermophilus* and *Escherichia coli*.
Eur. J. Biochem., **125**, pp 357-360.

BARSTOW D.A., SHARMAN A.F., ATKINSON T. et MINTON N.P. (1986a)

Cloning and complete nucleotide sequence of the *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase gene.
Gene, **46**, pp 37-45.

BARSTOW D.A., CLARKE A.R., CHIA W.N., WIGLEY D., SHARMAN A.F., HOLBROOK J.J., ATKINSON T. et MINTON N.P. (1986b)

Cloning, expression and complete nucleotide sequence of the *Bacillus stearothermophilus* L-lactate dehydrogenase gene.
Gene, **46**, pp 47-55.

BIESECKER G., HARRIS J.I., THIERRY J.C., WALKER J.E. et WONACOTT A.J. (1977)

Sequence and structure of D-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*.
Nature, **266**, pp 328-333.

BILLARD P. (1988)

Modification de la spécificité de coenzyme de la glyceraldehyde-3-phosphate déshydrogénase de *Bacillus stearothermophilus* par mutagenèse dirigée.
D.E.A. Université de Nancy 1.

BIRNBOIM H.C. et DOLY J. (1979)

A rapid alkaline extraction procedure for screening recombinant plasmid deoxyribonucleic acid.
Nucleic Acids Res., **7**, pp 1513-1523.

BOLIVAR F. et BACKMANN K. (1979)

Plasmids of *Escherichia coli* as cloning vectors.
Methods Enzymol., **68**, pp 245-285.

BOLIVAR F., RODRIGUEZ R.L., GREENE P.J., BETLACH M.C., HEYNECKER H.L. et BOYER H.W.W. (1977)

Construction and characterization of new cloning vehicles.
II. A multipurpose system.
Gene, **2**, pp 95-113.

BONEKAMP F., ANDERSEN H.D., CHRISTENSEN T. et JENSEN K.F. (1985)

Codon-defined ribosomal pausing in *Escherichia coli* detected by using the pyrE attenuator to probe the coupling between transcription and translation.
Nucleic Acids Res., **13**, pp 4113-4123.

BORGFORD T.J., BRAND N.J., GRAY T.E. et FERSCHT A.R. (1987)

The valyl-tRNA-synthetase from *Bacillus stearothermophilus* has considerable sequence homology with the isoleucyl-tRNA-synthetase from *Escherichia coli*.
Biochemistry, **26**, pp 2480-2486.

BRANLANT G. et BRANLANT C. (1985)

Nucleotide sequence of the *Escherichia coli* gap gene. Different evolutionary behaviour of the NAD⁺ binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase.
Eur. J. Biochem., **150**, pp 61-66.

BRANLANT G., FLESCH G. et BRANLANT C. (1983)

Molecular cloning of the glyceraldehyde-3-phosphate dehydrogenase of *Bacillus stearothermophilus* and *Escherichia coli*, and their expression in *Escherichia coli*.
Gene, **25**, pp 1-7.

BRINKMANN H., MARTINEZ P., QUIGLEY F., MARTIN W. et CERFF R. (1987)

Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize.
J. Mol. Evol., **26**, pp 320-328.

BROMBACH M., GUALERZI C.O., NAKAMURA Y. et PON C.L. (1986)

Molecular cloning and sequence of the *Bacillus stearothermophilus* translational initiation factor IF2 gene.
Mol. Gen. Genet., **203**, pp 97-102.

BUTLER J.S., SPRINGER M. et GRUNBERG-MANAGO M. (1987)

AUU-to-AUG mutation in the initiator codon of the translation initiation factor IF3 abolishes translational autocontrol of its own gene (infC) in vivo.
Proc. Natl. Acad. Sci. USA, **84**, pp 4022-4025.

CARTER P., BEDOUELLE H. et WINTER G. (1985)

Improved oligonucleotide site-directed mutagenesis using M13 vector.
Nucleic Acids Res., **13**, pp 4431-4443.

CERFF R. (1979)

Quaternary structure of higher plant glyceraldehyde-3-phosphate dehydrogenases.
Eur. J. Biochem., **94**, 243-247.

CERFF R. et CHAMBERS S.E. (1979)

Subunit structure of higher plant glyceraldehyde-3-phosphate dehydrogenases (EC 1.2.1.12 and EC 1.2.1.13).
J. Biol. Chem., **254**, pp 6094-6098.

CERFF R. et KLOPPSTECH K. (1982)

Structural diversity and differential light control of mRNAs coding for angiosperm glyceraldehyde-3-phosphate dehydrogenases.
Proc. Natl. Acad. Sci. USA, **79**, pp 7624-7628.

CHEN K.C.K., CHEN J.S. et JOHNSON J.L. (1986)

Structural features of multiple nifH-like sequences and very biased codon usage in nitrogenase genes of *Clostridium pasteurianum*.
J. Bacteriol., **166**, pp 162-172.

CONWAY A. et KOSHLAND Jr D.E. (1968)

Negative cooperativity in enzyme action. The binding of diphosphopyridine nucleotide to glyceraldehyde-3-phosphate dehydrogenase.
Biochemistry, **7**, pp 4011-4023.

CONWAY T, SEWELL G.W. et INGRAM L.O. (1987)

Glyceraldehyde-3-phosphate dehydrogenase gene from *Zymomonas mobilis* : cloning, sequencing and identification of promoter region.
J. Bacteriol., **169**, pp 5653-5662.

CONWAY T. et INGRAM L.O. (1988)

Phosphoglycerate kinase gene from *Zymomonas mobilis* : cloning, sequencing and localization within the gap operon.
J. Bacteriol., **170**, pp 1926-1933.

COOK B.A. et KOSHLAND Jr D.E. (1970)

Positive and negative cooperativity in yeast glyceraldehyde-3-phosphate dehydrogenase.
Biochemistry, **9**, pp 3337-3342.

CORI G.T., SLEIN M.W. et CORI C.F. (1945)

Isolation and crystallization of D-glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle.
J. Biol. Chem., **159**, p 565.

CUMMINS C.S. et JOHNSON J.L. (1971)

Taxonomy of the Clostridia. Wall composition and deoxyribonucleic acid homologies in *Clostridium butyricum* and other butyric acid Clostridia.
J. Gen. Microbiol., **67**, pp 33-46.

DALDAL F. et APPLEBAUM J. (1985)

Cloning and expression of *Clostridium pasteurianum* galactokinase gene in *Escherichia coli* K-12 and nucleotide sequence analysis of a region affecting the amount of the enzyme.
J. Mol. Biol., **186**, pp 533-545.

DOMDEY H., WIEBAUER K., KLAPTHOR H. et ARNOLD H.H. (1983)
 Sequence analysis of the cloned messenger RNA coding for glyceraldehyde-3-phosphate dehydrogenase from chicken heart muscle.
 Eur. J. Biochem., **131**, pp 129-135.

DOUTHWAITE S., CHRISTENSEN A. et GARETT R.A. (1983)
 High order structure in the 3'-minor domain of small subunit ribosomal RNAs from a Gram negative bacterium, a Gram positive bacterium and an eukaryote.
 J. Mol. Biol., **169**, pp 249-279.

DUGAICZYK A., HARON J.A., STONE E.M., DENNISON O.E., ROTHBLUM K.N. et SCHWARTZ R.J. (1983)
 Cloning and sequencing of a deoxyribonucleic acid copy of glyceraldehyde-3-phosphate dehydrogenase messenger ribonucleic acid isolated from chicken muscle.
 Biochemistry, **22**, pp 1605-1613.

EHRlich S.D. (1988)
 Stability of genetic information in *Bacillus subtilis*.
 Dans : Abstract book : "8th International Biotechnology Symposium, Paris"

FABRY S. et HENSEL R. (1987)
 Purification and characterization of D-glyceradehyde-3-phosphate dehydrogenase from the thermophilic archaebacterium *Methanothermus fervidus*.
 Eur. J. Biochem., **165**, pp 147-155.

FABRY S. et HENSEL R. (1988)
 Primary structure of glyceraldehyde-3-phosphate dehydrogenase deduced from the nucleotide sequence of the thermophilic archaebacterium *Methanothermus fervidus*.
 Gene, **64**, pp 189-197.

FEINBERG A.P. et VOGELSTEIN B. (1983)
 A technique for radiolabelling deoxyribonucleic acid restriction endonuclease fragments to high specific activity.
 Anal. Biochem., **132**, pp 6-13.

FEINBERG A.P. et VOGELSTEIN B. (1984)
 A technique for radiolabelling deoxyribonucleic acid restriction endonuclease fragments to high specific activity (Addendum).
 Anal. Biochem., **137**, pp 266-267.

FERDINAND W. (1964)
 The isolation and specific activity of rabbit muscle glyceraldehyde-3-phosphate dehydrogenase.
 Biochem. J., **92**, pp 578-585.

FREIER S.M., KIERZEK R., JAEGER J.A., SUGIMOTO N., CARUTHERS M.H., NEILSON T. et TURNER D.H. (1986)
 Improved free-energy parameters for predictions of RNA duplex stability.
 Proc. Natl. Acad. Sci. USA, **83**, pp 9373-9377.

FRENCH A. et CHANG S. (1987)
 Nucleotide sequence of the phosphofructokinase gene from *Bacillus stearothermophilus* and comparison with the homologous *Escherichia coli* gene.
 Gene, **54**, pp 65-71.

GAIT M.J. (1984)
 Oligonucleotide synthesis, a practical approach.
 IRL Press Limited, OXFORD.

- GRAVES M.C., MULLENBACH G.T. et RABINOWITZ J.C. (1985)**
Cloning and nucleotide sequence determination of *Clostridium pasteurianum* ferredoxin gene.
Proc. Natl. Acad. Sci. USA, **82**, pp 1653-1657.
- GRAVES M.C. et RABINOWITZ J.C. (1986)**
In vivo and in vitro transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in Gram⁺ organisms.
J. Biol. Chem., **261**, pp 11409-11415.
- HAGER P.W. et RABINOWITZ J.C. (1985)**
Translational specificity in *Bacillus subtilis*.
Dans : The molecular biology of the *Bacilli*, Vol. II, pp 1-32.
Dubnau D.A. (Ed). Academic Press, New York.
- HANAUER A. et MANDEL J.L. (1984)**
The glyceraldehyde-3-phosphate dehydrogenase gene family : structure of a human cDNA and of an X chromosome linked pseudogene ; amazing complexity of the gene family in mouse.
The EMBO J., **3**, pp 2627-2633.
- HARRIS J.I. et PERHAM R.N. (1968)**
Glyceraldehyde-3-phosphate dehydrogenase from pig muscle.
Nature, **219**, pp 1025-1028.
- HARRIS J.I. et WATERS M. (1976)**
Glyceraldehyde-3-phosphate dehydrogenase.
Dans : The Enzymes, Boyer P.D. (Ed). Academic Press, New York **13**, pp 1-48.
- HENSEL R., LAUMANN S., LANG J., HEUMANN H. et LOTTSPREICH F. (1987)**
Characterization of two glyceraldehyde-3-phosphate dehydrogenases from the extremely thermophilic archaebacterium *Thermoproteus tenax*.
Eur. J. Biochem., **170**, pp 325-333.
- HILLMAN J.D. et FRAENKEL D.G. (1975)**
Glyceraldehyde-3-phosphate dehydrogenase mutants of *Escherichia coli*.
J. Bacteriol., **122**, pp 1175-1179.
- HINTON S.M. et FREYER G. (1986)**
Cloning, expression and sequencing the molybdenum-pterin binding protein (mop) gene of *Clostridium pasteurianum* in *Escherichia coli*.
Nucleic Acids Res., **14**, pp 9371-9380.
- HOCKING J.D. et HARRIS J.I. (1980)**
D-glyceraldehyde-3-phosphate dehydrogenase amino-acid sequence of the enzyme from the extreme thermophile *Thermus aquaticus*.
Eur. J. Biochem., **108**, pp 567-579.
- HOLLAND J.P. et HOLLAND M.J. (1979a)**
Structural comparison of two nontandemly repeated yeast glyceraldehyde-3-phosphate dehydrogenase genes.
J. Biol. Chem., **254**, pp 2596-2604.
- HOLLAND M.J. et HOLLAND J.P. (1979b)**
Isolation and characterization of a gene coding for glyceraldehyde-3-phosphate dehydrogenase from *Saccharomyces cerevisiae*.
J. Biol. Chem., **254**, pp 5466-5474.

HOLLAND J.P., LABIENIC L., SWIMMER C. et HOLLAND M.J. (1983)
Homologous nucleotide sequence at the 5'-termini of messenger RNAs synthesised from the yeast enolase and glyceraldehyde-3-phosphate dehydrogenase gene families. The primary structure of a third yeast glyceraldehyde-3-phosphate dehydrogenase gene.
J. Biol. Chem., **258**, pp 5291-5299.

IGLESIAS A.A., SERRANO A., GUERRERO M.G. et LOSADA M. (1987)
Purification and properties of NADP-dependent non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase from the green alga *Chlamydomonas reinhardtii*.
Biochim. Biophys. Acta, **925**, pp 1-10.

IGLESIAS A.A. et LOSADA M. (1988)
Purification and kinetic and structural properties of spinach leaf NADP-dependent non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase.
Arch. Biochem. Biophys., **260**, pp 830-840.

IHARA H., SASAKI T., TSUBOI A., YAMAGATA H., TSUKAGOSHI N. et UDAKA S. (1985)

Complete nucleotide sequence of a thermophilic α -amylase gene : homology between prokaryotic and eukaryotic α -amylases at the active sites.
J. Biochem., **98**, pp 95-103.

IMURA T., UTATSU I. et TOH-E A. (1987)
Glyceraldehyde-3-phosphate dehydrogenase genes of *Zygosaccharomyces rouxii* : the source of a promoter for a host-vector system for *Z. rouxii*.
Agric. Biol. Chem., **51**, pp 1641-1647.

KAWAMOTO R.M. et CASWELL A.H. (1986)
Autophosphorylation of glyceraldehyde-3-phosphate dehydrogenase and phosphorylation of protein from skeletal muscle microsomes.
Biochemistry, **25**, pp 656-661.

KELLY G.J. and GIBBS M. (1973)
A mechanism for the indirect transfer of photosynthetically reduced nicotinamide adenine dinucleotide phosphate from chloroplasts to the cytoplasm.
Plant Physiol., **52**, pp 674-676.

KIRSCHNER K., GALLEGO E., SCHUSTER I. et GOODELL D. (1971)
Cooperative binding of nicotinamide-adenine dinucleotide to yeast glyceraldehyde-3-phosphate dehydrogenase.
J. Mol. Biol., **58**, pp 29-50.

KRAMER W., DRUTSA V., JENSEN H.W., KRAMER B., FLUGFELDER M.P. et FRITZ H.J. (1984)
The gapped duplex deoxyribonucleic acid approach to oligonucleotide directed mutation construction.
Nucleic Acids Res., **12**, pp 9441-9456.

LAEMMLI U.K. (1970)
Cleavage of structural proteins during the assembly of the head of the bacteriophage T4.
Nature, **227**, pp 680-686.

LESLIE A.G.W. et WONACOTT A.J. (1983)
Coenzyme binding in crystals of glyceraldehyde-3-phosphate dehydrogenase.
J. Mol. Biol., **165**, pp 375-391.

LESLIE A.G.W. et WONACOTT A.J. (1984)

Structural evidence for ligand-induced sequential conformational changes in glyceraldehyde-3-phosphate dehydrogenase.
J. Mol. Biol., **178**, pp 743-772.

LEWIN R. (1983)

How mammalian RNA returns to its genome.
Science, **219**, pp 1052-1054.

LOVENBERG W. et SOBEL B.E. (1965)

Rubredoxin : a new electron transfer protein from *Clostridium pasteurianum*.
Proc. Natl. Acad. Sci. USA, **54**, 193-199.

MACHATT M.A., EBEL J.P. et BRANLANT C. (1981)

The 3'-terminal region of bacterial 23S rRNA: structure and homology with the 3'-terminal region of eukaryotic 28S rRNA and with chloroplast 4.5S rRNA.
Nucleic Acids Res., **9**, pp 1533-1549.

Mc CLURE W.R. (1985)

Mechanism and control of transcription initiation in prokaryotes.
Annu. Rev. Biochem., **54**, pp 171-204.

Mc GOWAN R.E. et GIBBS M. (1974)

Comparative enzymology of the glyceraldehyde-3-phosphate dehydrogenase from *Pisum sativum*.
Plant physiol., **54**, pp 312-319.

Mc LAUGHLIN J.R., MURRAY C.L. et RABINOWITZ J.C. (1981)

Unique features in ribosome binding site sequence of the Gram positive *Staphylococcus aureus* β -lactamase gene.
J. Biol. Chem., **256**, pp 11283-11291.

MANIATIS T., FRITSCH E.F. et SAMBROOK J. (1982)

Dans : Molecular cloning : a laboratory manual.
 Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

MARGULIS L. (1970)

Dans : Origin of eukaryotic cells.
 Yale University Press. New Haven, Connecticut.

MARTIN W. et CERFF R. (1986)

Prokaryotic features of a nucleus encoded enzyme. cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase from mustard (*Sinapis alba*).
Eur. J. Biochem., **159**, pp 323-331.

MESSING J., GRONENBORN B., MULLER H.B. et HOFSCHEIDER P.H. (1977)

Filamentous coliphage M13 as a cloning vehicle : insertion of *HindIII* fragment of the lac regulatory region in M13 replicative form *in vitro*.
Proc. Natl. Acad. Sci. USA, **74**, pp 3642-3646.

MISSET O., VAN BEEUMEN J., LAMBEIR A.M., VAN DER MEER R. et OPPERDOES F.R. (1987)

Glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei*. Comparison of the glycosomal and cytosolic isoenzymes.
Eur. J. Biochem., **162**, pp 501-507.

- MONOT F., MARTIN J.R., PETITDEMANGE H. et GAY R. (1982)**
Acetone-butanol production by *Clostridium acetobutylicum* in a synthetic medium.
Appl. Environ. Microbiol., **44**, pp 1318-1324.
- MORAN C.P., LAND N., LE GRICE S.F., LEE G., STEPHENS M., SONENSHEIN A.L., PERO J. et LOSICK R. (1982)**
Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*.
Mol. Gen. Genet., **186**, pp 161-178.
- MORAS D., OLSEN K.W., SABESAN M.N., BUEHNER M., FORD G.C. et ROSSMANN M.G. (1975)**
Studies of asymmetry in the three-dimensional structure of lobster glyceraldehyde-3-phosphate dehydrogenase.
J. Biol. Chem., **250**, pp 9137-9162.
- MORI N., SINGER-SAM J. et RIGGS A.D. (1986)**
Evolutionary conservation of the substrate-binding cleft of phosphoglycerate kinases.
FEBS Lett., **204**, pp 313-317.
- MURTHY M.R.N., GARAVITO R.M., JOHNSON E.J. et ROSSMANN M.G. (1980)**
Structure of lobster apo D-glyceraldehyde-3-phosphate dehydrogenase at 3.0 Å resolution.
J. Mol. Biol., **138**, pp 859-872.
- NISHI T. et ITOH S. (1986)**
Enhancement of transcriptional activity of the *E.coli* trp promoter by upstream A+T-rich regions.
Gene, **44**, pp 29-36.
- PANABIÈRES F., PIECHACZYK M., RAINER B., DANI C., FORT P., RIAAD S., MARTY L., IMBACH J.L., JEANTEUR P. et BLANCHARD J.M. (1983)**
Complete nucleotide sequence of the messenger RNA coding for chicken muscle glyceraldehyde-3-phosphate dehydrogenase.
Biochem. Biophys. Res. Comm., **118**, pp 767-773.
- PAWLUK A., SCOPES R.K. et GRIFFITHS-SMITH K. (1986)**
Isolation and properties of the glycolytic enzyme from *Zymomonas mobilis*.
Biochem. J., **238**, pp 275-281.
- PIECHACZYK M., BLANCHARD J.M., RIAAD-EL-SABOUTZ S., DANI C., MARTY L. et JEANTEUR P. (1984)**
Unusual abundance of vertebrate glyceraldehyde-3-phosphate dehydrogenase pseudogenes.
Nature, **312**, pp 469-471.
- POLGAR L. (1975)**
Ion-pair formation as a source of enhanced reactivity of the essential thiol group of D-glyceraldehyde-3-phosphate dehydrogenase.
Eur. J. Biochem., **51**, pp 63-71.
- POOLMAN B., BOSMAN B., KIERS J. et KONINGS W.N. (1987)**
Control of glycolysis by glyceraldehyde-3-phosphate dehydrogenase in *Streptococcus cremoris* and *Streptococcus lactis*.
J. Bacteriol., **169**, pp 5887-5890.
- PRIBNOW D. (1975)**
Nucleotide sequence of a RNA polymerase binding site at an early T7 promoter.
Proc. Natl. Acad. Sci. USA, **72**, pp 784-789.

QUIGLEY F., MARTIN W.F. et CERFF R. (1988)

Intron conservation across the prokaryote / eukaryote boundary : structure of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize.
Proc. Natl. Acad. Sci. USA, **85**, pp 2672-2676.

RAFF R.A. et MAHLER H.R. (1972)

The nonsymbiotic origin of mitochondria.
Science, **177**, pp 575-582.

READ R.J., WIERENGA R.K., GROENDIJK H., HOL W.G.J., LAMBEIR A. et OPPERDOES F.R. (1987)

Preliminary crystallographic studies of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei*.
J. Mol. Biol., **194**, pp 573-575.

ROBINSON M., LILLEY R., LITTLE S., EMTAGE J.S., YARRANTOR G., STEPHENS P., MILLICAN A., EATON M. et HUMPHREYS G. (1984)

Codon usage can affect efficiency of translation of genes in *Escherichia coli*.
Nucleic Acids. Res., **12**, pp 6663-6671.

ROSSMAN M.G., ADAMS M.J., BUEHNER M., FORD G.C., HACKERT M.L., LILJAS A., RAO S.T., BANASZAK L.J., HILL E., TSERNOGLOU D. et WEBBS L. (1973)

Molecular symmetry axes and subunit interfaces in certain deshydrogenases.
J. Mol. Biol., **76**, pp 533-537.

ROSSMAN M.G., LILJAS A., BRANDEN C.I. et BANASZAK L.J. (1975)

Evolutionary and structural relationships among dehydrogenases.
Dans : The Enzymes, **11**, pp 61-102.
Boyer P.D. (Ed). Academic Press, New York.

SANGER F., NICKLEN S. et COULSON A.R. (1977)

DNA sequencing with chain terminating inhibition.
Proc. Natl. Acad. Sci. USA, **74**, pp 5463-5467.

SEEBURG P.H., NUSSLEIN C. et SCHALLER H. (1977)

Interaction of RNA polymerase with promoters from bacteriophage fd.
Eur. J. Biochem., **74**, pp 107-113.

SEGAL H.L. et BOYER P.D. (1953)

The role of sulfhydryl groups in the activity of D-glyceraldehyde-3-phosphate dehydrogenase.
J. Biol. Chem., **204**, pp 265-281.

SHARP P.A. (1983)

Conversion of RNA to DNA in mammals : *Alu* -like elements and pseudogenes.
Nature, **301**, pp 471-472.

SHARP P.M. et LI W.H. (1986)

Codon usage in regulatoty genes in *Escherichia coli* does not reflect selection for "rare" codons.
Nucleic Acids Res., **14**, pp 7737-7749.

SHIH M.C., LAZAR G. et GOODMAN H.M. (1986)

Evidence in favor of symbiotic origin of chloroplasts : primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenases.
Cell, **47**, pp 73-80.

SHIH M.C. et GOODMAN H.M. (1988)

Differential light regulated expression of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase in *Nicotiana tabacum*.

The EMBO J., **7**, pp 893-898.

SHINE J. et DALGARNO L. (1974)

The 3'-terminal sequence of *Escherichia coli* 16S rRNA. Complementary to non-sens triplets and ribosome binding sites.

Proc. Natl. Acad. Sci. USA, **71**, pp 1342-1346.

SHINE J. et DALGARNO L. (1975)

Determinant of cistron specificity in bacterial ribosomes.

Nature, **254**, pp 34-46.

SHONK C.E. et BOXER G.E. (1964)

Enzyme patterns in human tissues.

I. Methods for the determination of glycolytic enzymes.

Cancer Res., **24**, pp 709-721.

SKARZYNSKI T., MOODY P.C.E. et WONACOTT A.J. (1987)

Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1,8 Å resolution.

J. Mol. Biol., **193**, pp 171-187.

SKARZYNSKI T. et WONACOTT A.J. (1988)

Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*.

J. Mol. Biol., **203**, pp 1097-1118.

SOUKRI A. (1988)

Modification du site actif de la glyc eraldehyde-3-phosphate d shydrog nase par mutag n se dirig e :  tude du r le de la Cys 149 et de l'His 176 dans le m canisme enzymatique.

Th se d'Universit . Nancy I.

SOUTHERN E.M. (1975)

Detection of specific sequences among DNA fragments separated by gel electrophoresis.

J. Mol. Biol., **98**, pp 503-517.

SPASSKY A., BUSBY S. et BUC H. (1984)

On the action of the cyclic AMP - cyclic AMP receptor protein complex at the *Escherichia coli* lactose and galactose promoter regions.

The EMBO J., **3**, pp 43-50.

STALLCUP W.B. et KOSHLAND Jr D.E. (1973a)

Half of the sites reactivity and negative cooperativity. The case of yeast glyceraldehyde-3-phosphate dehydrogenase.

J. Mol. Biol., **80**, pp 41-62.

STALLCUP W.B. et KOSHLAND Jr D.E. (1973b)

Half of the sites reactivity in the catalytic mechanism of yeast glyceraldehyde-3-phosphate dehydrogenase.

J. Mol. Biol., **80**, pp 77-91.

STONE E.M., ROTHBLUM K.N. et SCHWARTZ R.J. (1985)

Intron-dependent evolution of chicken glyceraldehyde-3-phosphate dehydrogenase gene.

Nature, **313**, pp 498-500.

SULLIVAN D.T., CARROLL W.T., KANIK-ENNULAT C.L., HITTI Y.S., LOVETT J.A. et VON KALM L. (1985)

Glyceraldehyde-3-phosphate dehydrogenase from *Drosophila melanogaster*. Identification of two isozymic forms encoded by separated genes.
J. Biol. Chem., **260**, pp 4345-4350.

TAKAGI M., IMANAKA T. et AIBA S. (1985)

Nucleotide sequence and promoter region for the neutral protease gene from *Bacillus stearothermophilus*.
J. Bacteriol., **163**, pp 824-831.

TAKANAMI M., SUGIMOTO K., SUGISAKI H. et OKAMOTO T. (1976)

Sequence of promoter for coat protein gene of bacteriophage fd.
Nature, **260**, pp 297-309.

TOKUNAGA K., NAKAMURA Y., SAKATA K., FUJIMORI K., OHKUDO M., SAWADA K. et SAKIYAMA S. (1987)

Enhanced expression of a glyceraldehyde-3-phosphate dehydrogenase gene in human lung cancers.
Cancer Res., **47**, pp 5616-5619.

TSO J.Y., SUN X.H., KAO T.H., REECE K.S. et WU R. (1985a)

Isolation and characterization of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs : genomic complexity and molecular evolution of the gene.
Nucleic Acids Res., **13**, pp 2485-2502.

TSO J.Y., SUN X.H. et WU R. (1985b)

Structure of two unlinked *Drosophila melanogaster* glyceraldehyde-3-phosphate dehydrogenase genes.
Nucleic Acids. Res., **13**, pp 8220-8228.

VARENNE S. et LAZDUNSKI C. (1986)

Effects of distribution of unfavorable codons on the maximum rate of gene expression by an heterologous organism.
J. Theor. Biol., **120**, pp 99-110.

VIEIRA J. et MESSING J. (1982)

The pUC plasmids, an M13-mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers.
Gene, **19**, pp 259-268.

WALKER J.E., CARNE A.F., RUNSWICK M.J., BRIDGEN J. et HARRIS J.I. (1980a)

D-glyceraldehyde-3-phosphate dehydrogenase. Complete amino-acid sequence of the enzyme from *Bacillus stearothermophilus*.
Eur. J. Biochem., **108**, pp 549-565.

WALKER J.E., WONACOTT A.J. et HARRIS J.I. (1980b)

Heat stability of a tetrameric enzyme, D-glyceraldehyde-3-phosphate dehydrogenase.
Eur. J. Biochem., **108**, pp 581-586.

WANG S.Z., CHEN J.S. et JOHNSON J.L. (1987)

Nucleotide and deduced amino-acid sequences of nifD encoding the α -subunit of nitrogenase MoFe protein of nitrogenase Mo-Fe protein of *Clostridium pasteurianum*.
Nucleic Acids Res., **15**, pp 3935.

WAYE M.M.Y. et WINTER G. (1986)

A transcription terminator in the 5' non-coding region of the tyrosyl-tRNA synthetase gene from *Bacillus stearothermophilus*.

Eur. J. Biochem., **158**, pp 505-510.

WEEDEN N.F. (1981)

Genetic and biochemical implications of the endosymbiotic origin of the chloroplast.

J. Mol. Evol., **17**, pp 133-139.

WINTER G.P., KOCH G.L., HARTLEY B.S. and BARKER D.G. (1983)

The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*.

Eur. J. Biochem., **132**, pp 383-387.

**YARBROUGH P.O., HAYDEN M.A., DUNN L.A., VERMERSCH P.S.,
KLASS M.R. et HECHT R.M. (1987)**

The glyceraldehyde-3-phosphate dehydrogenase gene family in the nematode *Caenorhabditis elegans* : isolation and characterization of one of the genes.

Biochim. Biophys. Acta, **908**, pp 21-23.

RESUME

La glycéraldéhyde-3-phosphate déshydrogénase (GAPDH) est un modèle largement utilisé pour étudier les relations structure-fonction au sein d'une enzyme et l'évolution phylogénétique des espèces.

Nous avons cloné et séquencé les gènes de deux GAPDH bactériennes, celle de *C.pasteurianum* et celle de *B.stearothermophilus*. Nous avons ainsi pu approfondir les connaissances de la structure des gènes de ces deux bactéries Gram⁺. Nous avons étudié en détail les signaux de régulation de la transcription et de la traduction de chacun des deux gènes. Nous avons également constaté que, chez les deux bactéries étudiées, les gènes de la GAPDH et de la phosphoglycérate kinase sont associés en tandem dans une structure probablement de type opéron.

Nous avons réussi à surexprimer chez *E.coli* chacun des deux gènes de GAPDH, malgré un usage de codons peu favorable, surtout dans le cas de *C.pasteurianum*.

Enfin, nous avons étudié l'évolution phylogénétique des espèces en utilisant comme base la GAPDH. Cette étude a porté sur les 20 structures primaires différentes connues à ce jour. Elle a permis de confirmer la particularité de l'enzyme d'*E.coli* qui se rapproche davantage des enzymes cytosoliques d'eucaryotes que des autres enzymes bactériennes.

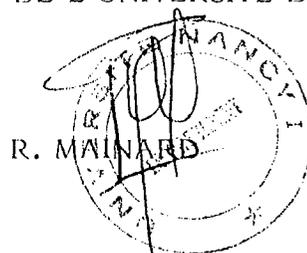
NOM DE L'ETUDIANT : OSTER Thierry

NATURE DE LA THESE : Doctorat de l'Université de Nancy I
en Biologie Moléculaire et Génie Génétique

VU, APPROUVE ET PERMIS D'IMPRIMER

NANCY, le 25 JAN 1999 n° 158

LE PRESIDENT DE L'UNIVERSITE DE NANCY I



UNIVERSITE DE NANCY I

20 JAN 89	000153
-----------	--------

COUPON ARRIVEE
SCOLARITE

MOTS-CLES

Glycéraldéhyde-3-phosphate déshydrogénase - *Clostridium pasteurianum* -
Bacillus stearothermophilus - séquences nucléotidiques -
expression hétérologue chez *Escherichia coli* - évolution phylogénétique

RESUME

La glycéraldéhyde-3-phosphate déshydrogénase (GAPDH) est un modèle largement utilisé pour étudier les relations structure-fonction au sein d'une enzyme et l'évolution phylogénétique des espèces.

Nous avons cloné et séquencé les gènes de deux GAPDH bactériennes, celle de *C.pasteurianum* et celle de *B.stearothermophilus*. Nous avons ainsi pu approfondir les connaissances de la structure des gènes de ces deux bactéries Gram⁺. Nous avons étudié en détail les signaux de régulation de la transcription et de la traduction de chacun des deux gènes. Nous avons également constaté que, chez les deux bactéries étudiées, les gènes de la GAPDH et de la phosphoglycérate kinase sont associés en tandem dans une structure probablement de type opéron.

Nous avons réussi à surexprimer chez *E.coli* chacun des deux gènes de GAPDH, malgré un usage de codons peu favorable, surtout dans le cas de *C.pasteurianum*.

Enfin, nous avons étudié l'évolution phylogénétique des espèces en utilisant comme base la GAPDH. Cette étude a porté sur les 20 structures primaires différentes connues à ce jour. Elle a permis de confirmer la particularité de l'enzyme d'*E.coli* qui se rapproche davantage des enzymes cytosoliques d'eucaryotes que des autres enzymes bactériennes.