



**HAL**  
open science

# Dynamique chromosomique et duplication génique chez *Streptomyces ambofaciens*

Virginie Roth

► **To cite this version:**

Virginie Roth. Dynamique chromosomique et duplication génique chez *Streptomyces ambofaciens*. Génétique. Université Henri Poincaré - Nancy 1, 2003. Français. NNT: 2003NAN10228. tel-01748193

**HAL Id: tel-01748193**

**<https://hal.univ-lorraine.fr/tel-01748193>**

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](mailto: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.cfcopies.com/V2/leg/leg_droi.php)

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



Thèse présentée pour l'obtention du titre de  
**DOCTEUR DE L'UNIVERSITÉ HENRI POINCARÉ, NANCY 1**  
en Génétique Moléculaire

par

**Virginie ROTH**

**DYNAMIQUE CHROMOSOMIQUE ET DUPLICATION GÉNIQUE**  
**CHEZ *STREPTOMYCES AMBOFACIENS***

Soutenue le 14 novembre 2003

Composition du jury :

*Rapporteurs :*

M. Philippe MAZODIER  
M. Serge POTIER

Chef de laboratoire, Institut Pasteur, Paris  
Professeur, Université Louis Pasteur, Strasbourg

*Examineurs :*

M. Stéphane AYMERICH  
M. Bernard DE CARIS  
M. Pierre LEBLOND (Dir. de thèse)

Directeur de Recherche, INRA, Paris-Grignon  
Professeur, Université Henri Poincaré, Nancy 1  
Professeur, Université Henri Poincaré, Nancy 1

BIBLIOTHEQUE SCIENCES NANCY 1



D 095 180830 4

## REMERCIEMENTS

Mes remerciements s'adressent à Monsieur le Professeur Bernard DECARIS.

« Je vous exprime toute ma gratitude et ma reconnaissance pour m'avoir offert la possibilité d'atteindre mon objectif en m'accueillant au sein de votre laboratoire ».

Ma reconnaissance s'adresse à Monsieur le Professeur Pierre LEBLOND.

« Je vous adresse tous mes remerciements pour m'avoir guidée dans mon travail, pour la qualité de votre encadrement et pour votre dévouement ».

Je remercie Monsieur le Docteur Philippe MAZODIER et Monsieur le Professeur Serge POTIER de m'avoir fait l'honneur et le plaisir d'être les rapporteurs de cette thèse.

Je remercie également Monsieur le Docteur Stéphane AYMERICH d'avoir accepté de juger ce travail.

Merci à Pierre pour ses conseils et sa disponibilité durant ces cinq dernières années.

Merci à Bertrand pour son soutien et son intérêt pour ces travaux.

Merci à toute l'équipe du laboratoire de Génétique et Microbiologie pour sa gentillesse et sa sympathie et particulièrement à Céline pour sa disponibilité et sa bonne humeur quotidienne.

Merci infiniment à Michel, Françoise, Céline et Matthieu qui ont toujours cru en moi.

Merci de tout cœur à Régis pour sa confiance et ses encouragements de tous les jours.

*Je dédie cette thèse à mes parents...*





A - PREMIERE PARTIE

LES REARRANGEMENTS GENOMIQUES : MECANISMES  
D'EVOLUTION RAPIDE 4

- CHAPITRE A-I -

IMPACT DES REARRANGEMENTS SUR L'EXPRESSION GENIQUE

A-I.1	œ Réarrangement et formation d'un gène	6
A-I.1.1	- Création d'une fonction au cours de la différenciation cellulaire	6
A-I.1.2	- Phénomènes de réarrangements génomiques créateurs de diversité	8
A-I.1.2.1	- Structure mosaïque des gènes des immunoglobulines et diversité des anticorps	8
A-I.1.2.2	- Variation des protéines de surface chez certaines bactéries pathogènes	9
A-I.2	œ Modulation de l'expression d'un gène par effet de position	11
A-I.2.1	- Interconversion du signe conjugal chez la levure <i>Saccharomyces cerevisiae</i>	11
A-I.2.2	- Mouvement de la fourche de réplication et expression génique chez <i>E. coli</i>	12
A-I.3	œ Réarrangement et création d'un nouveau cadre d'expression	13
A-I.3.1	- Modifications de signaux transcriptionnels	13
A-I.3.2	- Modifications de la traduction	14
A-I.4	œ Modulation de l'expression par modification du dosage génique	15
A-I.4.1	- Amplification de gènes de résistance aux drogues dans les cellules eucaryotes	15
A-I.4.2	- Dosage génique et configuration chromosomique chez les bactéries	16

- CHAPITRE A-II -

LA DUPLICATION GENIQUE

A-II.1	œ Mécanismes de la duplication de gènes	18
A-II.1.1	- La recombinaison homologue inégale	18
A-II.1.2	- La polyploïdisation	18

A-II.1.3 – La rétroposition _____	19
<b>A-II.2 ∞ Duplication et évolution _____</b>	<b>21</b>
A-II.2.1 - Le modèle Duplication-Complémentation-Dégénérescence _____	22
A-II.2.2 - Duplication et régulation de l'expression des gènes dupliqués _____	24
A-II.2.2.1 – Devenir des gènes dupliqués chez les eucaryotes _____	24
A-II.2.2.2 – Devenir des gènes dupliqués chez les bactéries _____	26
A-II.2.2.2.1 –Duplication de gènes aboutissant à la surproduction d'une protéine ____	26
A-II.2.2.2.2 –Duplication de gènes aboutissant à l'expression différentielle d'une protéine _____	27
A-II.2.2.2.2.1 – Les gènes <i>lysS</i> et <i>lysU</i> chez <i>E.coli</i> _____	27
A-II.2.2.2.2.2 – Les gènes <i>esp</i> chez <i>E. coli</i> _____	27
A-II.2.2.2.2.3 – Régulation différentielle de gènes dupliqués chez les <i>Streptomyces</i> _____	28
A-II.2.3 - Relation entre linéarité chromosomique, plasticité génomique et duplication génique _____	28
A-II.2.3.1 – Polymorphisme des régions subtélomériques _____	28
A-II.2.3.1.1. - Variation des régions subtélomériques eucaryotes _____	28
A-II.2.3.1.2. – Régions permissives aux réarrangements chez les bactéries _____	29
A-II.2.3.1.2.1 – Réarrangements dans les chromosomes circulaires bactériens _____	29
A-II.2.3.1.2.1 – Polymorphisme des régions subtélomériques des chromosomes linéaires bactériens _____	30
A-II.2.3.2 – Linéarité chromosomique et taux de redondance génomique _____	32
A-II.2.3.2.1 – Taux de redondance des génomes séquencés _____	32
A-II.2.3.2.2 – Localisation des duplicats ou familles de gènes _____	34
A-II.2.3.2.3 – Fonctions redondantes _____	36

### - CHAPITRE A-III -

#### L'INSTABILITE GENETIQUE AFFECTE L'EXPRESSION DU METABOLISME SECONDAIRE ET LA DIFFERENCIATION CELLULAIRE CHEZ LES *STREPTOMYCES*

<b>A-III.1 ∞ Réarrangements génomiques et polymorphisme intraspécifique chez les <i>Streptomyces</i> _____</b>	<b>41</b>
A-III.1.1 - Organisation génétique originale du chromosome _____	41
A-III.1.2 - L'instabilité génétique et la modulation de l'expression génique _____	42
A-III.1.2.1 - L'instabilité génétique est corrélée à la formation de réarrangements chromosomiques _____	42
A-III.1.2.2 - Impact de l'instabilité génomique sur l'expression génique _____	43
A-III.1.2.2.1 - Modulation du dosage génique _____	43
A-III.1.2.2.1.1 - Amplification d'ADN _____	43
A-III.1.2.2.1.2 - Variation intraspécifique de la taille des répétitions terminales _____	43
A-III.1.2.2.2 - Création d'un nouveau cadre d'expression des ORF <i>bas</i> _____	45
<b>A-III.2 ∞ La différenciation complexe des <i>Streptomyces</i> et son réseau de régulation _____</b>	<b>46</b>

## B - SECONDE PARTIE :

# REGULATION DE L'INITIATION DE LA TRANSCRIPTION CHEZ LES BACTERIES \_\_\_\_\_ 48

### - CHAPITRE B-I -

#### LE MECANISME D'INITIATION DE LA TRANSCRIPTION

B-I.1	∞ Le facteur sigma : sous-unité de l'ARN polymérase ADN dépendante _____	53
B-I.1.1	– Les acteurs : l'ARN polymérase ADN dépendante _____	53
B-I.1.1.1	– La forme apoenzyme de l'ARN polymérase ADN dépendante _____	53
B-I.1.1.2	– La forme holoenzyme de l'ARN polymérase ADN dépendante et les facteurs sigma principaux _____	54
B-I.1.1.2.1	– Les facteurs sigma principaux chez les bactéries _____	54
B-I.1.1.2.2	– La structure des facteurs sigma de type $\sigma^{70}$ _____	54
B-I.1.2	– Le mécanisme de l'initiation de la transcription _____	56
B-I.2	∞ Les séquences promotrices reconnues par les facteurs sigma de type $\sigma^{70}$ _____	58

### - CHAPITRE B-II -

#### REGULATION TRANSCRIPTIONNELLE ET FACTEURS SIGMA ALTERNATIFS

B-II.1	∞ Présentation des facteurs sigma alternatifs _____	60
B-II.2	∞ Les facteurs sigma alternatifs et la complexité de réponse à un stress environnemental d'un organisme _____	61
B-II.2.1	– Les facteurs sigma alternatifs chez <i>Escherichia coli</i> _____	61
B-II.2.1.1	– Le facteur $\sigma^{32}$ et la réponse au choc thermique _____	62
B-II.2.1.2	– Le facteur $\sigma^{24}$ ou $\sigma^E$ et le choc thermique prolongé _____	63
B-II.2.1.3	– Le facteur $\sigma^{28}$ ou $\sigma^F$ et la mobilité _____	63
B-II.2.1.4	– Le facteur $\sigma^{38}$ ou $\sigma^S$ et l'entrée en phase stationnaire _____	63
B-II.2.1.5	– Le facteur $\sigma^{19}$ ou protéine FecI et le transport du citrate ferrique _____	64
B-II.2.1.6	– Le facteur $\sigma^{54}$ ou $\sigma^N$ et le métabolisme de l'azote _____	64
B-II.2.2	– Les facteurs sigma alternatifs chez <i>Bacillus subtilis</i> _____	64
B-II.2.2.1	– La sporulation chez <i>Bacillus subtilis</i> _____	64
B-II.2.2.2	– Le facteur $\sigma^B$ et la réponse générale au stress chez <i>Bacillus subtilis</i> _____	67

B-II.2.3 – Les facteurs sigma alternatifs chez <i>Streptomyces</i> _____	68
B-II.2.3.1 – La différenciation et la réponse aux stress chez <i>Streptomyces coelicolor</i> _____	68
B-II.2.3.1.1 - Les facteurs $\sigma$ WhiG et $\sigma^F$ chez <i>Streptomyces coelicolor</i> _____	68
B-II.2.3.1.2 - Le facteur $\sigma$ BldN chez <i>Streptomyces coelicolor</i> _____	69
B-II.2.3.1.3 - Le facteur $\sigma^{H1}$ chez <i>Streptomyces coelicolor</i> _____	69
B-II.2.3.1.4 - Le facteur $\sigma^B$ chez <i>Streptomyces coelicolor</i> _____	70
B-II.2.3.1.5 - Le facteur $\sigma^R$ chez <i>Streptomyces coelicolor</i> _____	70
B-II.2.3.2.3 - Le facteur $\sigma$ sigG chez <i>Streptomyces coelicolor</i> _____	70
B-II.2.3.3 – Les (autres) facteurs $\sigma$ de type ECF chez <i>Streptomyces coelicolor</i> _____	71
B-II.2.3.3.1 - Le facteur $\sigma^F$ chez <i>Streptomyces coelicolor</i> _____	71
B-II.2.3.3.2 - Le facteur $\sigma^I$ chez <i>Streptomyces coelicolor</i> _____	72
B-II.2.3.4 - Chez les autres <i>Streptomyces</i> : illustration de la diversité des facteurs sigma _____	72

## MATERIEL ET METHODES

73

### - CHAPITRE I -

#### MANIPULATION DES SOUCHES BACTERIENNES ET VECTEURS DE CLONAGE

<b>I.1</b> Souches bactériennes, vecteurs et conditions de croissance _____	75
I.1.1 - Souches bactériennes _____	75
I.1.2 - Vecteurs de clonage _____	76
I.1.3 - Milieux et conditions de culture _____	79
I.1.4 - Courbes de croissance _____	81
I.1.4.1 - Courbes de croissance par mesure de la concentration en ADN _____	81
I.1.4.2 - Courbes de croissance par mesure de la densité optique _____	83
<b>I.2</b> Transformation de souches bactériennes _____	84
I.2.1 - Electroporation d' <i>E. coli</i> _____	84
I.2.1.1 - Introduction des cosmides dans <i>E. coli</i> BW25113/pIJ790 _____	84
I.2.1.2 - Construction des cosmides modifiés _____	85
I.2.2 - Conjugaison intraspécifique <i>E. coli</i> S17 / <i>E. coli</i> SURE™. _____	85
I.2.3 - Conjugaison intergénérique <i>E. coli</i> / <i>S. ambofaciens</i> _____	85
I.2.4 - Transformation de protoplastes de <i>Streptomyces</i> _____	87
<b>I.3</b> Observation des colonies bactériennes à la recherche d'un phénotype _____	88
I.3.1 - Observations morphologiques _____	88
I.3.2 - Traitements aux agents oxydants chimiques _____	88



I.3.3 - Traitement aux rayonnements ultraviolets (UV)	88
I.3.4 - Détermination de l'activité antibactérienne	89

**- CHAPITRE II -  
MANIPULATION DES ACIDES NUCLEIQUES**

<b>II.1 ☞ Préparation des acides nucléiques</b>	<b>91</b>
II.1.1 - Préparation d'ADN, digestion enzymatique et électrophorèse	91
II.1.1.1 - Extraction et purification d'ADN	91
II.1.1.2 - Digestions par les endonucléases et électrophorèse d'ADN	91
II.1.1.3 - Ligation	92
II.1.2 - Préparation et dosage d'ARN	93
<b>II.2 ☞ Les techniques utilisant les acides nucléiques</b>	<b>94</b>
II.2.1 - Hybridation ADN - ADN	94
II.2.2 - Séquençage automatique d'ADN	94
II.2.3 - Transcription inverse (RT) et amplification d'ADN (PCR)	94
II.2.3.1 - Difficultés de la méthode dans le cas des bactéries	94
II.2.3.2 - Choix d'un témoin interne pour une approche qualitative : le gène <i>recA</i>	96
II.2.3.3 - Les milieux réactionnels de RT-PCR	97
II.2.3.3.1 - La transcription inverse des ARN en ADNc	97
II.2.3.3.2 - L'amplification d'ADN par polymérisation en chaîne (PCR)	98
II.2.4 - Extension d'amorces	101
II.2.5 - Cartographie à la nucléase S1	102
<b>II.3 ☞ Analyse des séquences d'ADN</b>	<b>103</b>
II.3.1 - Détection de phases ouvertes de lectures	103
II.3.2 - Analyse de la redondance d'un génome par le logiciel BLASTCLUST	103

**- CHAPITRE III -  
LA PCR QUANTITATIVE**

<b>III.1 ☞ Principe de la PCR quantitative</b>	<b>105</b>
III.1.1 - Les bases de la PCR quantitative en temps réel	105
III.1.2 - La technologie de détection basée sur le SYBR Green	105
III.1.3 - La notion de cycle seuil ou Ct (threshold cycle)	107
III.1.4 - Efficacité et spécificité de la réaction	108

III.1.4.1 - Efficacité de la PCR et courbe standard	108
III.1.4.2 - Spécificité de la PCR et courbe de fusion	109
III.2 ☞ Un exemple d'analyse de résultats de PCR quantitative	111

☞ RESULTATS ☞	115
---------------	-----

A - PREMIÈRE PARTIE :
-----------------------

DYNAMIQUE DE LA DUPLICATION CHROMOSOMIQUE CHEZ <i>STREPTOMYCES AMBOFACIENS</i>	119
---	-----

- CHAPITRE A-I -

MISE EN EVIDENCE DE FUSIONS DE CHROMOSOMES PARTIELLEMENT DELETES

A-I.1 ☞ Structure chromosomique fusionnée des mutants <i>S. ambofaciens</i> NSA27 et NSA65	121
A-I.1.1 – Le mutant <i>S. ambofaciens</i> NSA27	121
A-I.1.2 – Le mutant <i>S. ambofaciens</i> NSA65	125
A-I.2 ☞ Analyse des séquences impliquées dans le réarrangement chez <i>S. ambofaciens</i> NSA27	128

- CHAPITRE A-II -

LA FUSION CHROMOSOMIQUE : UN ETAT MUTATEUR ?

A-II.1 ☞ Fusion de chromosomes et instabilité chromosomique	131
A-II.1.1 – Instabilité chromosomique des mutants <i>S. ambofaciens</i> NSA27 et NSA65	131
A-II.1.2 - Un cycle d'instabilité chromosomique	133
A-II.2 ☞ Fusion de chromosomes et instabilité phénotypique	135
A-II.2.1 – Instabilité phénotypique des mutants <i>S. ambofaciens</i> NSA27 et NSA65 et de leurs descendants	135
A-II.2.2 – Corrélation entre état fusionné et défaut de sporulation	136
A-II.2.2.1 – Analyse du contenu en ADN des spores par coloration au DAPI	136

A-II.2.2.2 – Analyse de l’ultrastructure des spores par microscopie électronique_____	137
A-II.2.2.3 – Analyse du taux de germination_____	139

**B - SECONDE PARTIE :**

<b>UN EVENEMENT DE DUPLICATION GENIQUE DANS LES REGIONS TERMINALES DU CHROMOSOME DE <i>STREPTOMYCES AMBOFACIENS</i></b> _____	<b>140</b>
---	------------

**- CHAPITRE B-I -  
CARACTERISATION DES GENES *HAS***

<b>B-I.1</b> ☞ Les gènes <i>hasR</i> et <i>hasL</i> chez la souche sauvage <i>Streptomyces ambofaciens</i> DSM40697_____	<b>142</b>
B-I.1.1 – Les facteurs $\sigma^{\text{Has}}$ _____	142
B-I.1.2 – L’environnement génétique des gènes <i>has</i> _____	146
<b>B-I.2</b> ☞ Les gènes chimériques <i>has135</i> et <i>has855</i> chez les mutants <i>Streptomyces ambofaciens</i> NSAH135 et NSA855_____	<b>151</b>

**- CHAPITRE B-II -  
LA DUPLICATION DES GENES *HAS* CHEZ *STREPTOMYCES AMBOFACIENS* EST RECENTE**

<b>B-II.1</b> ☞ Conservation de l’architecture de la région « core » du chromosome de <i>S. ambofaciens</i> _____	<b>154</b>
B-II.1.1 – Détermination des séquences des ITS de <i>S. ambofaciens</i> _____	154
B-II.1.2 – Mise en évidence de blocs constants et variables dans les ITS de <i>Streptomyces</i> _	155
B-II.1.3 – Comparaison des séquences ITS_____	156
B-II.1.4 – Prédiction des structures secondaires du repliement des séquences ITS _____	157
<b>B-II.2</b> ☞ Conservation de la duplication des gènes <i>has</i> chez les trois souches de <i>Streptomyces ambofaciens</i> _____	<b>159</b>
<b>B-II.3</b> ☞ Les gènes <i>has</i> au sein du genre <i>Streptomyces</i> _____	<b>163</b>

- CHAPITRE B-III -  
ANALYSE TRANSCRIPTIONNELLE DES GENES DUPLIQUES *HASR* ET *HASL*

<b>B-III.1</b>	<b>œ Analyse qualitative de la transcription des gènes <i>has</i> par RT-PCR</b>	<b>167</b>
B-III.1.1	– Expression des gènes <i>hasR</i> et <i>hasL</i> chez la souche sauvage <i>S. ambfaciens</i> DSM40697	167
B-III.1.2	– Expression des gènes <i>hasR</i> et <i>hasL</i> et des chimères <i>has855</i> et <i>has135</i> au cours de la croissance	169
<b>B-III.2</b>	<b>œ Identification des transcrits <i>has</i> et analyse de la régulation de leur expression</b>	<b>173</b>
<hr/>		
B-III.2.1	– Identification des sites d’initiation de la transcription et des promoteurs des gènes <i>has</i>	173
B-III.2.1.1	– Identification des +1 des gènes <i>has</i> par une approche d’extension d’amorces	173
B-III.2.1.2	– Identification des sites +1 des gènes <i>has</i> par une approche de cartographie à la nucléase S1 et analyse des séquences promotrices	175
B-III.2.1.2.1	- Les transcrits du gène <i>hasR</i>	175
B-III.2.1.2.1.1	– Le transcrit initié au site <i>hasRp1</i> et son promoteur de type $\sigma^B$	175
B-III.2.1.2.1.2	– Le transcrit initié au site <i>hasRp2</i> et son promoteur de type $\sigma^B$	177
B-III.2.1.2.1.3	– Les promoteurs putatifs du gène <i>hasR</i> reconnus par <i>WbiG</i>	178
B-III.2.1.2.2	- Les transcrits du gène <i>hasL</i>	180
B-III.2.1.2.2.1	– Le transcrit initié au site <i>hasL.p1</i> et son promoteur de type $\sigma^B$	181
B-III.2.1.2.2.2	– Le transcrit initié au site <i>hasL.p2</i> et son promoteur de type $\sigma^{HNR}$	182
B-III.2.2	– Régulation de l’expression des transcrits <i>has</i> au cours de la croissance	183
B-III.2.2.1	– Identification de la régulation de la transcription des gènes <i>has</i> au cours de la croissance par cartographie à la nucléase S1	183
B-III.2.2.2	– Quantification des niveaux d’induction de la transcription des gènes <i>has</i> au cours de la croissance par PCR quantitative	184
B-III.2.2.2.1	– Principes	184
B-III.2.2.2.1.1	– Principe de la quantification de chaque transcrit <i>has</i>	184
B-III.2.2.2.1.2	– Traitement mathématique des données	186
B-III.2.2.2.2	– Quantification de la transcription du gène <i>hasR</i>	187
B-III.2.2.2.2.1	– Evolution de la quantité de chaque transcrit du gène <i>hasR</i> au cours de la croissance	187
B-III.2.2.2.2.2	– Abondance relative des transcrits initiés aux sites <i>hasRp1</i> et <i>hasRp2</i>	188
B-III.2.2.2.3	– Quantification de la transcription du gène <i>hasL</i>	188
B-III.2.2.2.3.1	– Evolution de la quantité de chaque transcrit du gène <i>hasL</i> au cours de la croissance	188
B-III.2.2.2.3.2	– Abondance relative des transcrits initiés aux sites <i>hasL.p1</i> et <i>hasL.p2</i>	189
B-III.2.2.2.4	– Expression relative des gènes <i>hasR</i> et <i>hasL</i> au cours de la croissance	191
<b>B-III.3</b>	<b>œ Analyse de l’expression spatio-temporelle des promoteurs <i>has</i></b>	<b>192</b>
III.3.1	– Utilisation du gène rapporteur codant la résistance à la kanamycine	192

B-III.3.1.1 - Clonage des régions promotrices des gènes <i>hasR</i> et <i>hasL</i> _____	192
B-III.3.1.2 - Cartographie de restriction du plasmide pNSL9 _____	194
B-III.3.1.2.1 - Cartographie de restriction <i>BstBI</i> du plasmide pNSL9 _____	194
B-III.3.1.2.2 - Cartographie de restriction <i>ApaI</i> du vecteur recombiné pNSL9 _____	195
B-III.3.1.3 - Détection d'une force promotrice en amont de <i>hasL</i> _____	196
B-III.3.1.3.1 - Expression de la résistance à la kanamycine chez <i>S. lividans</i> NSL9 en croissance sur milieu solide _____	196
B-III.3.1.3.2 - Expression de la résistance à la kanamycine chez <i>S. lividans</i> NSL9 au cours de la croissance en milieu liquide _____	197
B-III.3.2 – Utilisation du gène rapporteur codant L'EGFP _____	198
B-III.3.2.1 - Clonage des régions promotrices des gènes <i>hasL</i> et <i>hasR</i> dans le vecteur pIJ8630 _____	198
B-III.3.2.2 - Cartographie de restriction des plasmides pIJR et pIJL _____	200
B-III.3.2.2.1 – Les vecteurs recombinants pIJL _____	200
B-III.3.2.2.2 – Les vecteurs recombinants pIJR _____	202
B-III.3.2.3 - Analyse de la fluorescence chez <i>S. ambofaciens</i> _____	204

#### - CHAPITRE B-IV -

#### CONSTRUCTION DE MUTANTS INTERROMPUS POUR LE(S) GENE(S) *HAS*

<b>B-IV.1</b> ☞ Stratégie de remplacement d'une ORF au locus chromosomique chez <i>Streptomyces</i> _____	<b>207</b>
B-IV.1.1 – Choix des amorces et fabrication du produit de PCR (étape 1) _____	209
B-IV.1.2 – Construction de l'allèle modifié sur le cosmide recombinant (étape 2) _____	209
B-IV.1.3 – Transfert conjugatif du cosmide modifié d' <i>E. coli</i> vers <i>Streptomyces</i> et sélection des recombinants _____	210
<b>B-IV.2</b> ☞ Construction des simples mutants $\Delta$ <i>hasR</i> et $\Delta$ <i>hasL</i> _____	<b>211</b>
B-IV.2.1 – Introduction des cosmides 14C4 et 25E1 dans la souche <i>E. coli</i> BW25113/pIJ790 _____	211
B-IV.2.2 – Amplification des cassettes <i>Apra</i> _____	212
B-IV.2.3 – Remplacement des ORF <i>hasL</i> et <i>hasR</i> par la cassette <i>Apra</i> dans les cosmides recombinants _____	212
B-IV.2.3.1 – Electroporation et sélection _____	212
B-IV.2.3.2 – Analyse des transformants _____	213
B-IV.2.4 – Remplacement des ORF <i>has</i> aux loci chromosomiques chez <i>S. ambofaciens</i> DSM40697 _____	214
B-IV.2.4.1 – Introduction des cosmides modifiés dans la souche <i>E. coli</i> ET12567/pUZ8002 _____	214



B-IV.2.4.2 – Conjugaisons intergénériques entre <i>E. coli</i> ET12567/pUZ8002/14C4 $\Delta$ <i>hasL</i> ::Apra ou 25E1 $\Delta$ <i>hasR</i> ::Apra et <i>S. ambofaciens</i> DSM40697	215
B-IV.2.4.3 – Mise en évidence de l'échange allélique aux loci <i>has</i> chromosomiques	215
B-IV.2.4.3.1 – Vérification par hybridation	215
B-IV.2.4.3.2 – Vérification par PCR	217
B-IV.2.4.4 – Vérification de l'intégrité de la structure chromosomique chez les simples mutants <i>S. ambofaciens</i> $\Delta$ <i>has</i> ::Apra	219
<b>B-IV.3 <math>\propto</math> Construction de doubles mutants <math>\Delta</math><i>hasR</i> <math>\Delta</math><i>hasL</i></b>	<b>220</b>
B-IV.3.1 – Amplification des cassettes Spec et remplacement dirigé des loci <i>has</i> sur les cosmides recombinants 14C4 et 25E1	220
B-IV.3.2 – Remplacement du locus <i>has</i> sauvage par la cassette Spec chez les simples mutants $\Delta$ <i>hasL</i> ::Apra et $\Delta$ <i>hasR</i> ::Apra	222
B-IV.3.3 – Mise en évidence du remplacement des deux ORF <i>has</i> aux loci chromosomiques	222
B-IV.3.3.1 – Vérification par hybridation	222
B-IV.3.3.2 – Vérification par PCR	223
IV.3.4 – Intégrité des régions terminales du chromosome des doubles mutants	224
<b>B-IV.4 <math>\propto</math> Construction de complémentants des simples mutants <math>\Delta</math><i>has</i></b>	<b>226</b>
B-IV.4.1 – Stratégie	226
B-IV.4.2 – Clonage des régions codantes des gènes <i>hasL</i> et <i>hasR</i> dans le vecteur pSET $\Omega$	227
B-IV.4.3 – Introduction chez <i>streptomyces</i>	227

- CHAPITRE B-V -

ANALYSE PHENOTYPIQUE DES MUTANTS  $\Delta$ *HAS*

<b>B-V.1 <math>\propto</math> Les facteurs <math>\sigma^{\text{Has}}</math> et la différenciation cellulaire</b>	<b>230</b>
B-V.1.1 – Croissance sur milieux gélosés	230
B-V.1.2 – Activité antibactérienne	232
<b>B-V.2 <math>\propto</math> Les facteurs <math>\sigma^{\text{Has}}</math> et la réponse aux stress</b>	<b>233</b>
B-V.2.1 – Recherche de phénotypes associés aux stress	233
B-V.2.1.1 – Influence de la température sur le phénotype colonial	233
B-V.2.1.2 – Stress acide-alcalin	234
B-V.2.1.3 – Stress osmotique	234
B-V.2.1.4 – Stress UV	235
B-V.2.1.5 – Stress oxydant	236
B-V.2.2 – Analyse transcriptionnelle des gènes <i>has</i> en conditions de stress	238

B-V.2.2.1 – La transcription des gènes <i>hasR</i> et <i>hasL</i> est induite dans des conditions de stress _____	238
B-V.2.2.2 – Implication d'autres facteurs $\sigma$ dans l'induction de la transcription des gènes <i>has</i> dans des conditions de stress _____	239

**- CHAPITRE B-VI -**

**REGULATION DE L'EXPRESSION DES GENES *HASR* ET *HASL* PAR LES FACTEURS  $\sigma^{HASR}$  ET  $\sigma^{HASL}$  IN VIVO**

B-VI.1 $\propto$ Principe de l'analyse _____	242
B-VI.2 $\propto$ Analyse de l'impact de l'interruption des gènes <i>has</i> sur l'expression de chaque transcrit <i>has</i> _____	243
B-VI.2.1 – Effet de l'Interruption d'un gène <i>has</i> sur sa propre transcription _____	247
B-VI.2.1.1 – Interruption d'un gène <i>has</i> et diminution de son niveau de transcription_	247
B-VI.2.1.2 – Interruption d'un gène <i>has</i> et augmentation de son niveau de transcription _____	248
B-VI.2.2 – Effet de l'Interruption d'un gène <i>has</i> sur la transcription de l'autre copie <i>has</i> _	249
B-VI.2.3 – Effet de l'Interruption des deux gènes <i>has</i> _____	251
B-VI.2.4 – Complémentation fonctionnelle _____	252

<b><math>\propto</math> DISCUSSION <math>\propto</math> _____</b>	<b>254</b>
---	------------

<b>A - PREMIERE PARTIE :</b>
------------------------------

<b>COMPLEXITE DE LA REGULATION DE L'EXPRESSION GENIQUE CHEZ LES <i>STREPTOMYCES</i> _____</b>	<b>255</b>
---	------------

**- CHAPITRE A-I -**

**LES DIVERS NIVEAUX DE REGULATION DE L'EXPRESSION DES GENES CODANT LES FACTEURS SIGMA CHEZ LES *STREPTOMYCES***

A-I.1 $\propto$ Le contrôle transcriptionnel des gènes <i>has</i> _____	257
A-I.1.1 – La transcription des gènes <i>has</i> est sous contrôle de promoteurs multiples _____	257
A-I.1.2 – Les facteurs sigma Has et la répression transcriptionnelle _____	258

A-I.1.3 – Modulation de la régulation croisée des gènes <i>has</i> en fonction du stade de croissance _____	258
<b>A-I.2 ☞ L'activité des facteurs sigma Has est-elle régulée par des mécanismes post-traductionnels ? _____</b>	<b>261</b>
A-I.2.1 - Synthèse d'un facteur sigma sous forme de propeptide (pro-sigma) _____	261
A-I.2.2 - La protéolyse, mécanisme de contrôle de la différenciation cellulaire _____	262
A-I.2.3 - Contrôle de l'activité par des inhibiteurs (anti-facteur sigma) _____	263

- CHAPITRE A-II -  
LA REPONSE AU STRESS CHEZ LES *STREPTOMYCES*

A-II.1 ☞ La redondance des systèmes de régulation transcriptionnelle _____	266
A-II.2 ☞ Les facteurs sigma : un lien entre différenciation morphologique et réponse au stress _____	267

**B - SECONDE PARTIE :**

**CREATION DE DIVERSITE GENETIQUE PAR REARRANGEMENT**

<b>D'ADN _____</b>	<b>268</b>
--------------------	------------

- CHAPITRE B-I -  
POLYMORPHISME DES EXTREMITES DES REPLICONS LINEAIRES

B-I.1 ☞ Distribution de la linéarité chromosomique chez les bactéries _____	270
<b>B-I.2 ☞ Réplicons linéaires et réarrangements chromosomiques _____</b>	<b>272</b>
B-I.2.1 – La linéarité chromosomique confère-t-elle un avantage ? _____	272
B-I.2.1.1 – Régiospécificité des réarrangements chez les <i>Streptomyces</i> _____	272
B-I.2.1.2 – La linéarité chromosomique favorise-t-elle l'instabilité chromosomique ? _____	273
B-I.2.2 - Mécanismes moléculaires des réarrangements chromosomiques _____	274
B-I.2.3 - réarrangements chromosomiques et évolution rapide des génomes _____	275
B-I.2.3.1 - Recombinaison subtélomérique et acquisition d'information génétique _____	276
B-I.2.3.1.1 - Transfert horizontal d'information génétique chez les <i>Streptomyces</i> _____	276
B-I.2.3.1.2 – Mécanismes de transfert d'information génétique aux extrémités des chromosomes des <i>Streptomyces</i> _____	279

B-I.2.3.1.3 – Interactions plasmide/chromosome et acquisition de la linéarité chromosomique	281
B-I.2.3.1.2.1 - Interaction entre un plasmide linéaire et un chromosome circulaire ancestral chez <i>S. coelicolor</i> et <i>B. burgdorferi</i>	281
B-I.2.3.1.2.2 - Acquisition de fonctions chromosomiques sur un plasmide linéaire ancestral chez <i>Agrobacterium tumefaciens</i>	282
B-I.2.3.2 – Les extrémités terminales inversées des chromosomes de <i>Streptomyces</i> ont-elles un rôle ?	283
B-I.2.3.2.1 - Implication de l'instabilité dans l'évolution rapide des extrémités des <i>Streptomyces</i>	283
B-I.2.3.2.2 – Les TIR : un substrat pour la recombinaison terminale ?	284

**- CHAPITRE B-II -**  
**DUPLICATIONS CHROMOSOMIQUES ET EVOLUTION**

<b>B-II.1</b> ∞ La dynamique des génomes	287
B-II.1.1 – Les <i>Streptomyces</i> : modèle d'étude de l'évolution des génomes	287
B-II.1.2 – Evolution des génomes : transfert horizontal et duplication génique	288
<b>B-II.2</b> ∞ Le devenir des gènes dupliqués <i>hasR</i> et <i>hasL</i> au cours de l'évolution	291
B-II.2.1 – Un mécanisme d'évolution concertée à l'origine de la conservation de la duplication des gènes <i>has</i> ?	291
B-II.2.2 – L'avantage évolutif de la duplication des gènes <i>has</i>	292

**C - TROISIEME PARTIE :**

<b>PERSPECTIVES</b>	<b>295</b>
---------------------	------------

<b>C-I</b> ∞ Valeurs adaptatives des duplications terminales	296
<b>C-II</b> ∞ Contrôle de la réponse au stress chez les <i>Streptomyces</i>	298
C-II.1 – Contrôle de l'expression des gènes <i>has</i>	298
C-II.1.1 – Contrôle la transcription des gènes <i>has</i>	298
C-II.1.2 – Les anti-facteurs $\sigma$ contrôlant l'activité des facteurs $\sigma^{\text{Has}}$	299
C-II.2 – Le rôle du régulon $\sigma^{\text{Has}}$	299
C-II.2.1 – Recherche de phénotype chez des mutants <i>has</i>	299
C-II.2.2 – Identification des gènes cibles des facteurs $\sigma^{\text{Has}}$	300

❧ BIBLIOGRAPHIE ❧ \_\_\_\_\_ 302

❧ PUBLICATIONS ❧ \_\_\_\_\_ 328

Publication n°1 \_\_\_\_\_ 329

Publication n°2 \_\_\_\_\_ 340

Publication n°3 \_\_\_\_\_ 356





L'introduction de ce manuscrit permettra d'exposer la problématique du travail en présentant, dans un premier temps, les réarrangements et leurs implications évolutives, notamment dans le cas de la duplication génique. Dans un second temps, un état des connaissances sur les systèmes de régulation de l'initiation de la transcription chez les bactéries modèles *Escherichia coli* et *Bacillus subtilis* mais également chez les *Streptomyces*, permettra d'aborder le contexte de cette étude de gènes dupliqués codant des facteurs  $\sigma$ .

Le chapitre « Matériel et Méthodes » comprend notamment l'exposé de l'approche de PCR quantitative qui a été abondamment utilisée dans les approches fonctionnelles.

Ensuite les résultats concernant la caractérisation d'une structure chromosomique conférant un état de mutateur chromosomique, générateur d'une grande diversité de duplications chromosomiques, seront résumés dans une première partie. L'ensemble des données est accessible dans les annexes (voir section « Publications »). Puis, l'analyse fonctionnelle des gènes *has* est détaillée dans une seconde partie avec notamment la construction de mutants et une analyse transcriptionnelle fine des gènes *has* par RT-PCR quantitative (article en préparation pour *Journal of Bacteriology*).

Enfin, la complexité des systèmes de régulation de l'expression chez les *Streptomyces*, notamment en réponse aux variations des conditions environnementales sera discutée. Puis, le chromosome de *S. ambifaciens* est proposé comme modèle d'étude des mécanismes d'évolution par duplications chromosomiques et d'intégration d'information exogène aux extrémités du chromosome linéaire.

# ❧ INTRODUCTION ❧

## A - PREMIERE PARTIE

### LES REARRANGEMENTS GENOMIQUES : MECANISMES

#### D'EVOLUTION RAPIDE \_\_\_\_\_ 4

#### - CHAPITRE A-I -

#### IMPACT DES REARRANGEMENTS SUR L'EXPRESSION GENIQUE

<b>A-I.1</b> ∞ Réarrangement et formation d'un gène _____	6
A-I.1.1 - Création d'une fonction au cours de la différenciation cellulaire _____	6
A-I.1.2 - Phénomènes de réarrangements génomiques créateurs de diversité _____	8
A-I.1.2.1 - Structure mosaïque des gènes des immunoglobulines et diversité des anticorps	8
A-I.1.2.2 - Variation des protéines de surface chez certaines bactéries pathogènes _____	9
<b>A-I.2</b> ∞ Modulation de l'expression d'un gène par effet de position _____	11
A-I.2.1 - Interconversion du signe conjugal chez la levure <i>Saccharomyces cerevisiae</i> _____	11
A-I.2.2 - Mouvement de la fourche de réplication et expression génique chez <i>E. coli</i> _____	12
<b>A-I.3</b> ∞ Réarrangement et création d'un nouveau cadre d'expression _____	13
A-I.3.1 - Modifications de signaux transcriptionnels _____	13
A-I.3.2 - Modifications de la traduction _____	14
<b>A-I.4</b> ∞ Modulation de l'expression par modification du dosage génique _____	15
A-I.4.1 - Amplification de gènes de résistance aux drogues dans les cellules eucaryotes _____	15
A-I.4.2 - Dosage génique et configuration chromosomique chez les bactéries _____	16

#### - CHAPITRE A-II -

#### LA DUPLICATION GENIQUE

<b>A-II.1</b> ∞ Mécanismes de la duplication de gènes _____	18
A-II.1.1 – La recombinaison homologue inégale _____	18
A-II.1.2 – La polyploïdisation _____	18
A-II.1.3 – La rétroposition _____	19
<b>A-II.2</b> ∞ Duplication et évolution _____	21
A-II.2.1 - Le modèle Duplication-Complémentation-Dégénérescence _____	22

A-II.2.2 - Duplication et régulation de l'expression des gènes dupliqués	24
A-II.2.2.1 – Devenir des gènes dupliqués chez les eucaryotes	24
A-II.2.2.2 – Devenir des gènes dupliqués chez les bactéries	26
A-II.2.2.2.1 –Duplication de gènes aboutissant à la surproduction d'une protéine	26
A-II.2.2.2.2 –Duplication de gènes aboutissant à l'expression différentielle d'une protéine	27
A-II.2.2.2.2.1 – Les gènes <i>lysS</i> et <i>lysU</i> chez <i>E.coli</i>	27
A-II.2.2.2.2.2 – Les gènes <i>csp</i> chez <i>E. coli</i>	27
A-II.2.2.2.2.3 – Régulation différentielle de gènes dupliqués chez les <i>Streptomyces</i>	28
A-II.2.3 - Relation entre linéarité chromosomique, plasticité génomique et duplication génique	28
A-II.2.3.1 – Polymorphisme des régions subtélomériques	28
A-II.2.3.1.1. - Variation des régions subtélomériques eucaryotes	28
A-II.2.3.1.2. – Régions permissives aux réarrangements chez les bactéries	29
A-II.2.3.1.2.1 – Réarrangements dans les chromosomes circulaires bactériens	29
A-II.2.3.1.2.1 – Polymorphisme des régions subtélomériques des chromosomes linéaires bactériens	30
A-II.2.3.2 – Linéarité chromosomique et taux de redondance génomique	32
A-II.2.3.2.1 – Taux de redondance des génomes séquencés	32
A-II.2.3.2.2 – Localisation des duplicats ou familles de gènes	34
A-II.2.3.2.3 – Fonctions redondantes	36

### - CHAPITRE A-III -

#### L'INSTABILITE GENETIQUE AFFECTE L'EXPRESSION DU METABOLISME SECONDAIRE ET LA DIFFERENCIATION CELLULAIRE CHEZ LES *STREPTOMYCES*

<b>A-III.1</b> ⚡ Réarrangements génomiques et polymorphisme intraspécifique chez les <i>Streptomyces</i>	41
A-III.1.1 - Organisation génétique originale du chromosome	41
A-III.1.2 - L'instabilité génétique et la modulation de l'expression génique	42
A-III.1.2.1 - L'instabilité génétique est corrélée à la formation de réarrangements chromosomiques	42
A-III.1.2.2 - Impact de l'instabilité génomique sur l'expression génique	43
A-III.1.2.2.1 - Modulation du dosage génique	43
A-III.1.2.2.1.1 - Amplification d'ADN	43
A-III.1.2.2.1.2 - Variation intraspécifique de la taille des répétitions terminales	43
A-III.1.2.2.2 - Création d'un nouveau cadre d'expression des ORF <i>has</i>	45
<b>A-III.2</b> ⚡ La différenciation complexe des <i>Streptomyces</i> et son réseau de régulation	46

## B - SECONDE PARTIE :

### REGULATION DE L'INITIATION DE LA TRANSCRIPTION CHEZ LES BACTERIES \_\_\_\_\_ 48

#### - CHAPITRE B-I -

##### LE MECANISME D'INITIATION DE LA TRANSCRIPTION

B-I.1	œ Le facteur sigma : sous-unité de l'ARN polymérase ADN dépendante _____	53
B-I.1.1	– Les acteurs : l'ARN polymérase ADN dépendante _____	53
B-I.1.1.1	– La forme apoenzyme de l'ARN polymérase ADN dépendante _____	53
B-I.1.1.2	– La forme holoenzyme de l'ARN polymérase ADN dépendante et les facteurs sigma principaux _____	54
B-I.1.1.2.1	– Les facteurs sigma principaux chez les bactéries _____	54
B-I.1.1.2.2	– La structure des facteurs sigma de type $\sigma^{70}$ _____	54
B-I.1.2	– Le mécanisme de l'initiation de la transcription _____	56
B-I.2	œ Les séquences promotrices reconnus par les facteurs sigma de type $\sigma^{70}$ _____	58

#### - CHAPITRE B-II -

##### REGULATION TRANSCRIPTIONNELLE ET FACTEURS SIGMA ALTERNATIFS

B-II.1	œ Présentation des facteurs sigma alternatifs _____	60
B-II.2	œ Les facteurs sigma alternatifs et la complexité de réponse à un stress environnemental d'un organisme _____	61
B-II.2.1	– Les facteurs sigma alternatifs chez <i>Escherichia coli</i> _____	61
B-II.2.1.1	– Le facteur $\sigma^{32}$ et la réponse au choc thermique _____	62
B-II.2.1.2	– Le facteur $\sigma^{24}$ ou $\sigma^E$ et le choc thermique prolongé _____	63
B-II.2.1.3	– Le facteur $\sigma^{28}$ ou $\sigma^F$ et la mobilité _____	63
B-II.2.1.4	– Le facteur $\sigma^{38}$ ou $\sigma^S$ et l'entrée en phase stationnaire _____	63
B-II.2.1.5	– Le facteur $\sigma^{19}$ ou protéine FecI et le transport du citrate ferrique _____	64
B-II.2.1.6	– Le facteur $\sigma^{54}$ ou $\sigma^N$ et le métabolisme de l'azote _____	64
B-II.2.2	– Les facteurs sigma alternatifs chez <i>Bacillus subtilis</i> _____	64
B-II.2.2.1	– La sporulation chez <i>Bacillus subtilis</i> _____	64
B-II.2.2.2	– Le facteur $\sigma^B$ et la réponse générale au stress chez <i>Bacillus subtilis</i> _____	67
B-II.2.3	– Les facteurs sigma alternatifs chez <i>Streptomyces</i> _____	68

B-II.2.3.1 – La différenciation et la réponse aux stress chez <i>Streptomyces coelicolor</i>	68
B-II.2.3.1.1 - Les facteurs $\sigma$ WhiG et $\sigma^F$ chez <i>Streptomyces coelicolor</i>	68
B-II.2.3.1.2 - Le facteur $\sigma$ BldN chez <i>Streptomyces coelicolor</i>	69
B-II.2.3.1.3 - Le facteur $\sigma^H$ chez <i>Streptomyces coelicolor</i>	69
B-II.2.3.1.4 - Le facteur $\sigma^B$ chez <i>Streptomyces coelicolor</i>	70
B-II.2.3.1.5 - Le facteur $\sigma^R$ chez <i>Streptomyces coelicolor</i>	70
B-II.2.3.2.3 - Le facteur $\sigma$ sigG chez <i>Streptomyces coelicolor</i>	70
B-II.2.3.3 – Les (autres) facteurs $\sigma$ de type ECF chez <i>Streptomyces coelicolor</i>	71
B-II.2.3.3.1 - Le facteur $\sigma^E$ chez <i>Streptomyces coelicolor</i>	71
B-II.2.3.3.2 - Le facteur $\sigma^U$ chez <i>Streptomyces coelicolor</i>	72
B-II.2.3.4 - Chez les autres <i>Streptomyces</i> : illustration de la diversité des facteurs sigma	72

A - PREMIERE PARTIE :

LES REARRANGEMENTS  
GENOMIQUES : MECANISMES  
D'EVOLUTION RAPIDE



**- CHAPITRE A-I -**

**IMPACT DES REARRANGEMENTS SUR  
L'EXPRESSION GENIQUE**

De nombreuses études ont démontré que les réarrangements génomiques ont un impact prépondérant au cours de l'évolution des eucaryotes comme des procaryotes.

Les mécanismes moléculaires de la plasticité des génomes sont variés : les recombinaisons « homologues » et « homéologues » sont impliquées, de même qu'un ensemble de mécanismes regroupés sous le terme "illégitime", incluant, entre autres, la recombinaison site-spécifique et les erreurs de réplication.

Les réarrangements « programmés » et les réarrangements stochastiques « mutationnels » sont souvent distingués. Les premiers affectent des fonctions et des lignées cellulaires spécifiques et interviennent dans une forte proportion de la population concernée. Le plus souvent, ils ne sont pas transmis à la descendance. Ils sont en cela compris comme des mécanismes de contrôle de l'expression génique au même titre que les contrôles transcriptionnels et traductionnels. Les phénomènes de réarrangements stochastiques affectent une certaine proportion de la population à chaque génération, et créent une diversité génétique, substrat des mécanismes de l'évolution.

Malgré ces distinctions, au regard de l'évolution, ces réarrangements génomiques ont été sélectionnés, les uns pour l'avantage conféré par la capacité de différenciation, et les autres pour la création d'un polymorphisme génétique.

## A-I.1 $\omega$ REARRANGEMENT ET FORMATION D'UN GÈNE

Chez de nombreux micro-organismes et lignées cellulaires eucaryotes, la différenciation cellulaire s'accompagne de remaniements génétiques établissant des fonctions spécifiques. Les fonctions concernées de même que les mécanismes moléculaires impliqués sont variés et seuls quelques exemples seront ici présentés.

### A-I.1.1 - CREATION D'UNE FONCTION AU COURS DE LA DIFFÉRENCIATION CELLULAIRE

Les réarrangements d'ADN accompagnant la sporulation (phénomène induit à l'entrée de la phase stationnaire) chez la bactérie modèle des Gram+ à haut pourcentage en bases G et C, *Bacillus subtilis*, et la différenciation de cellules fixatrices d'azote chez la cyanobactérie *Anabaena species*, sont très similaires (Haselkorn, 1992). Ces deux exemples peuvent constituer les archétypes des réarrangements programmés chez les bactéries.

Chez *B. subtilis*, un événement de recombinaison site-spécifique permet la formation du gène *sigK*, chimère des gènes *spoIVCB* et *spoIIIC*, codant le facteur  $\sigma^K$  sous forme d'un propeptide (Figure I-1 ; (Haselkorn, 1992)). Ce réarrangement consiste en l'excision d'un segment d'ADN de 32 kb

(appelé élément skin, (Popham and Stragier, 1992)) catalysée par la recombinase site-spécifique codée par l'élément skin (Kunkel *et al.*, 1990). La formation de ce réarrangement est nécessaire à l'établissement des étapes ultérieures de la sporulation (maturation de l'endospore). La sporulation intervient en réponse à une carence nutritionnelle de manière synchrone dans plus de 80% de la population cellulaire. Le génome de la spore n'est pas affecté, et le réarrangement n'est donc pas transmis à la descendance.

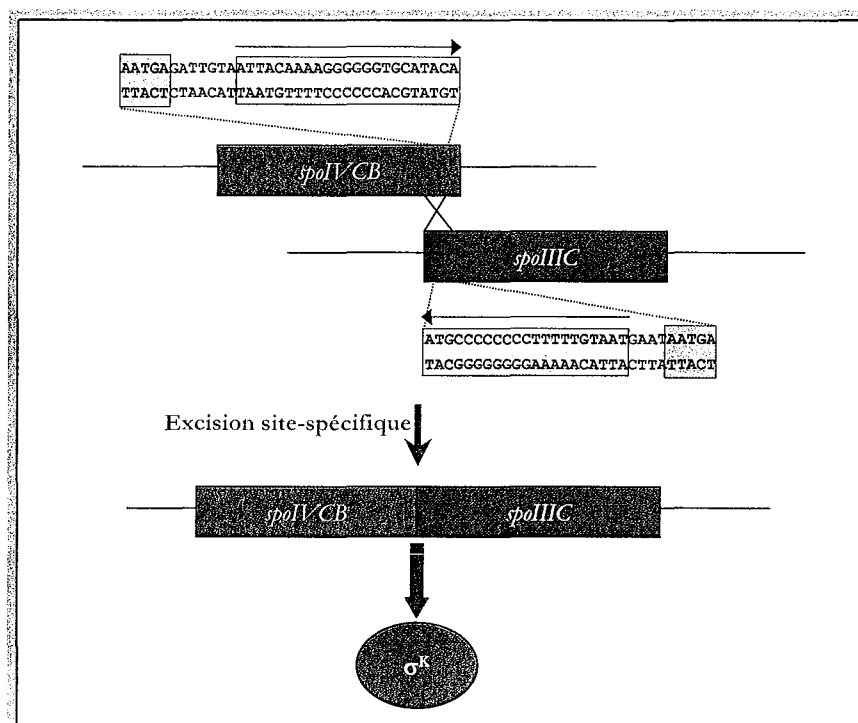


Figure I-1 : Illustration de la création d'une fonction par un réarrangement : recombinaison site-spécifique permettant la formation du gène *sigK*, chimère entre les gènes *spoIVCB* et *spoIIIC*, codant le facteur  $\sigma^K$ .  
Selon (Haselkorn, 1992).

Chez la cyanobactérie *Anabaena sp.*, deux réarrangements d'ADN accompagnent la différenciation de cellules végétatives en cellules fixant l'azote, appelées hétérocystes (Haselkorn, 1992). Les bactéries *Anabaena* croissent en filaments pluricellulaires, et environ une cellule sur dix subit la différenciation. Les réarrangements consistent en l'excision d'éléments d'ADN inclus dans les gènes *nifD* et *fdxN* (éléments de 11 kb et 55 kb respectivement, Figure I-2). Un événement de recombinaison site-spécifique, catalysé par les recombinases XisA et XisF codées par les éléments excisibles, intervient entre de courtes répétitions flanquant l'élément (Lammers *et al.*, 1986). Le gène *nifD*, appartenant à l'opéron *nifHDK*, code l'un des 3 polypeptides composant le complexe nitrogénase, responsable de la réduction de l'azote atmosphérique en ammonium. L'excision permet l'expression de *nifD* et donc la formation du complexe et la différenciation en hétérocyste

fixateur d'azote. Le gène *fdxN* code, quant à lui, une protéine "ferredoxine-like" exprimée uniquement dans les hétérocystes. Son rôle n'est cependant pas élucidé.

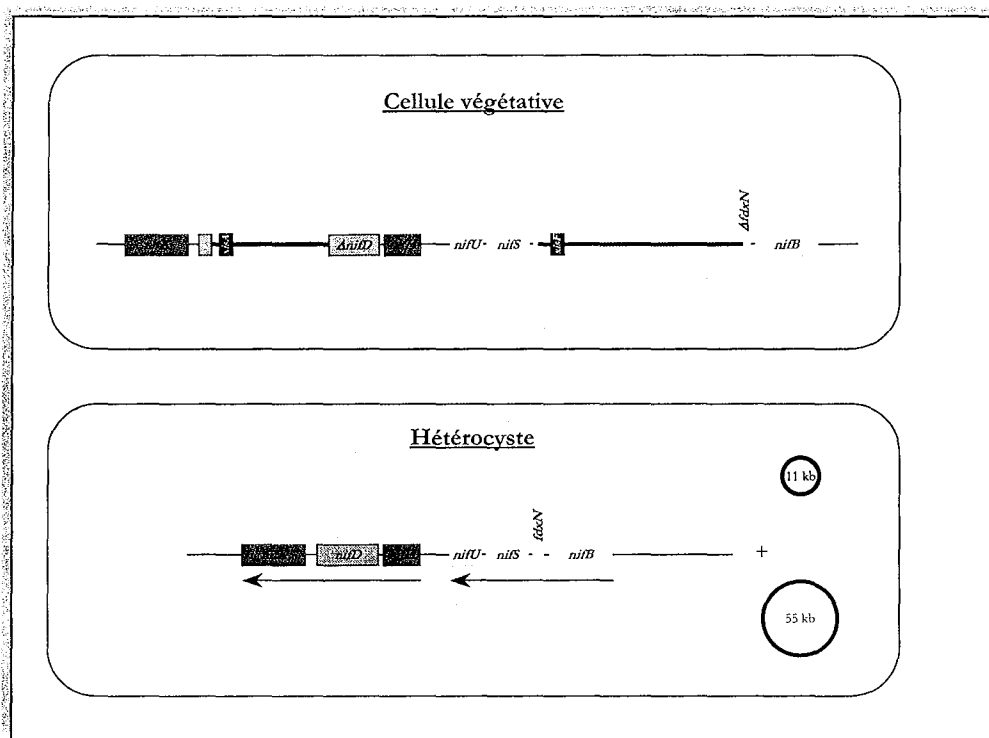


Figure I-2: Illustration de la création d'une fonction par un réarrangement: excision d'éléments d'ADN inclus dans les gènes *nifD* et *fdxN* chez la cyanobactérie *Anabaena sp.* Selon (Haselkorn, 1992).

## A-I.1.2 - PHENOMENES DE REARRANGEMENTS GENOMIQUES CREATEURS DE DIVERSITE

### A-I.1.2.1 - STRUCTURE MOSAÏQUE DES GENES DES IMMUNOGLOBULINES ET DIVERSITE DES ANTICORPS

Chez les mammifères, les lymphocytes B sont les cellules spécialisées dans la production d'anticorps spécifiques intervenant dans la réponse immunitaire. Les anticorps ou immunoglobulines sont constitués de deux chaînes polypeptidiques lourdes (ou H) et de deux chaînes légères (ou L) ; chaque chaîne est composée d'une région constante et d'une région variable.

La variabilité des anticorps résulte de 3 phénomènes majeurs affectant les gènes structuraux : la recombinaison VDJ, l'hypermutabilité somatique et la diversité jonctionnelle. En effet, les gènes codant chacune des régions variables résultent de la juxtaposition de différents segments d'ADN : segments V ("variable"), D ("diversity") et J ("joining ") pour les chaînes H, segments V et J seulement pour les chaînes L.

L'extrême variabilité des anticorps découle de l'existence de nombreux segments V, D et J regroupés sur un même chromosome. Par exemple chez la souris, pour la région variable de la chaîne lourde, il existe entre 200 et 2000 segments V, 12 segments D et 4 segments J (Blackwell and Alt, 1989). Une recombinase site-spécifique réarrange les segments d'ADN par un mécanisme appelé " recombinaison VDJ " (Hozumi and Tonegawa, 1976; Tonegawa, 1976). A ce processus de fusion des segments variables se surajoute le phénomène de diversité jonctionnelle. Aux jonctions des segments des chaînes lourdes (D-J et V-D-J) est catalysée l'addition de nucléotides par une terminal-déoxynucléotidyl-transférase (TdT). Enfin, un mécanisme, appelé hypermutabilité somatique, implique l'introduction de mutations à hautes fréquences à l'intérieur des domaines variables des chaînes lourdes et légères d'un anticorps exprimé.

#### A-I.1.2.2 - VARIATION DES PROTEINES DE SURFACE CHEZ CERTAINES BACTERIES PATHOGENES

Chez la bactérie pathogène *Neisseria gonorrhoeae*, agent de la gonococcie, une importante variabilité est observée au niveau des gènes codant les pili, nécessaires à l'adhérence aux cellules épithéliales hôtes. De nombreux gènes codant les pili sont présents sur le chromosome de *Neisseria* (Hass *et al.*, 1992). La plupart des gènes, au nombre de 17, sont silencieux (*pilS*) et sont dépourvus de région 5' codante. Des segments d'ADN des loci *pilS*, appelés mini-cassettes, remplacent des régions spécifiques des gènes *pilE*, loci exprimés, par un mécanisme de recombinaison homologue. Ces mini-cassettes sont bornées par des séquences conservées, substrats des événements de recombinaison. Un autre mécanisme impliquant la recombinaison homologue entraîne la formation de délétions de segments de gènes *pilE* et leur remplacement par des séquences issues des gènes *pilS* (Hill, 1992). Plus d'un million de combinaisons de séquences de gènes codant les pili peut être généré. Ainsi, la variabilité des profils d'expression de pili permet d'échapper à la réponse immunitaire de l'hôte.

Chez *Borrelia hermsii*, agent des fièvres récurrentes, la spécificité sérotypique est déterminée par des lipoprotéines membranaires variables (Restrepo *et al.*, 1992). De plus, au polymorphisme des gènes *vmp* s'ajoute la capacité de recombinaison de ces gènes entre eux. Ainsi, des gènes chimériques sont formés par la fusion de deux gènes *vmp* tronqués grâce à un événement de recombinaison homologue (Kitten *et al.*, 1993). La création de chimères est également observée chez *Borrelia burgdorferi*, responsable de la maladie de Lyme (Rosa *et al.*, 1992).

Par ailleurs, les gènes codant les antigènes de surface de nombreux pathogènes eucaryotiques comme procaryotiques renferment fréquemment des motifs répétés en tandem. Les domaines répétés de la protéine codée seraient des domaines de fixation à la surface des cellules hôtes (Wren, 1991). Des variations de la taille de ces gènes, résultant de la duplication ou de la délétion

d'une ou plusieurs copies de ces unités répétées, sont observées à haute fréquence. C'est notamment le cas des lipoprotéines variables (Vlps) présentes à la surface des cellules de *Mycoplasma hyorhinis* (agent d'infections chroniques), où des variations de taille se produisent à des fréquences de  $10^{-2}$  à  $10^{-4}$  par modification du nombre de répétitions internes (répétitions de 36 ou 39 paires de bases, (Citti and Wise, 1995). De même, chez *Pseudomonas aeruginosa* (agent d'infections opportunistes) des variations de taille du gène *algP*, impliqué dans la régulation des exopolysaccharides mucoïdes de la surface bactérienne, créent un polymorphisme important (Deretic and Konyecsni, 1990). Cette variation résulterait du glissement de la fourche de réplication (slipped-strand mispairing, SSM) sur les séquences répétées présentes à l'intérieur du gène *algP*. Ce mécanisme serait donc RecA-indépendant.

## **A-I.2 $\alpha$ MODULATION DE L'EXPRESSION D'UN GÈNE PAR EFFET DE POSITION**

La localisation, l'organisation et l'orientation transcriptionnelle d'un gène sur le chromosome ont été sélectionnées au cours de l'évolution. La modification du contexte d'un gène va provoquer le changement du profil d'expression de ce dernier : c'est l'effet de position.

L'effet de position chez les eucaryotes a été initialement décrit par (Muller, 1930) chez *Drosophila melanogaster*. Il démontre alors l'effet inhibiteur de la proximité d'une région hétérochromatique sur l'expression du gène *white* en étudiant les réarrangements des chromosomes suite à un traitement aux rayons X.

### **A-I.2.1 - INTERCONVERSION DU SIGNE CONJUGAL CHEZ LA LEVURE *SACCHAROMYCES CEREVISIAE***

Plusieurs loci distincts sont requis pour l'interconversion du signe conjugal (mating type) chez la levure *S. cerevisiae* (Klar, 1987) : le locus *MAT* (d'allèle  $a$  ou  $\alpha$ ), exprimé, et conférant à la levure son type sexuel, et les loci silencieux *HML $\alpha$*  et *HMR $a$*  (Figure I-3.). Ces gènes ( $a$  ou  $\alpha$ ) codent des régulateurs transcriptionnels de fonctions requises pour la conjugaison, notamment une phéromone. Enfin, le locus *HO* code une endonucléase introduisant des césures au locus *MAT*. Ces coupures initient l'événement de conversion génique de l'un des loci *HM* silencieux vers le locus *MAT* où il sera exprimé. Ce phénomène fait apparaître au sein d'une population homothallique, caractérisée par un type conjugal instable, des cellules de signes conjugaux opposés favorisant ainsi la sexualité et donc le brassage du patrimoine génétique.

L'inhibition de l'expression des gènes aux loci *HML* et *HMR* résulte de la fixation des protéines SIR, régulateurs d'information silencieuse sur des motifs ("silencers") en *cis*. Ces protéines maintiennent une structure analogue à l'hétérochromatine aux loci *HM* par contraste avec la structure des loci exprimés (Ansari and Gartenberg, 1999).

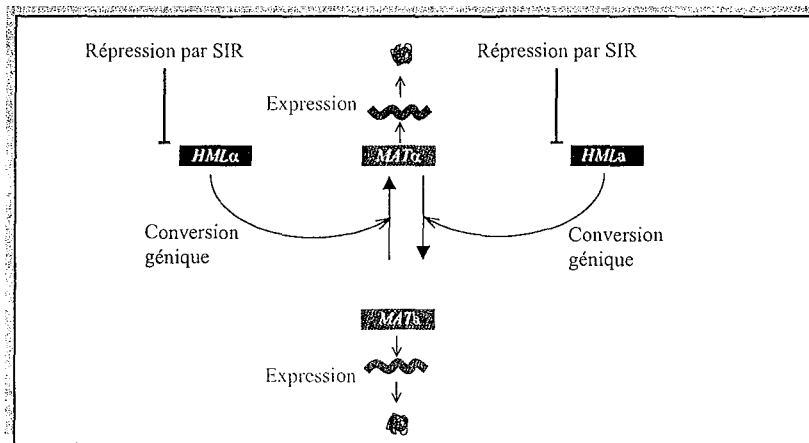


Figure I-3 : Illustration de réarrangements entraînant un changement des profils d'expression. Interconversion du signe conjugal chez la levure.

Cette carte représente les 3 loci du chromosome III, impliqués dans l'interconversion du signe conjugal. SIR signifie Régulateurs d'Informations Silencieuses.

Les loci exprimés sont représentés en vert, les loci silencieux en bleu, : ARN . : protéines

### A-I.2.2 - MOUVEMENT DE LA FOURCHE DE RÉPLICATION ET EXPRESSION GÉNÉTIQUE CHEZ *E. COLI*

Lorsque le déplacement du complexe de transcription converge avec celui de la fourche de réplication, la collision entraîne le décrochage de l'ARN polymérase, ainsi qu'une diminution de la célérité de la réplication (Brewer, 1988). Ainsi, 92% des gènes codant les composants de la machinerie traductionnelle (ARNr et protéines ribosomiques) sont transcrits dans le sens de déplacement de la fourche de réplication. La situation est similaire pour 88% des gènes impliqués dans la réplication et la transcription. Ces gènes ont des niveaux d'expression parmi les plus élevés (Brewer, 1988).

Ainsi, les réarrangements génomiques ayant pour conséquence une modification de l'orientation de gènes fortement transcrits vont d'une part augmenter le temps de réplication du chromosome (French, 1992) et d'autre part moduler le niveau d'expression des gènes fortement transcrits. Ces nouveaux génotypes pourront alors être sélectionnés.



## A-I.3 ∞ REARRANGEMENT ET CREATION D'UN NOUVEAU CADRE D'EXPRESSION

### A-I.3.1 - MODIFICATIONS DE SIGNAUX TRANSCRIPTIONNELS

La modification du cadre de transcription des gènes par les réarrangements génomiques est fréquente. Les éléments transposables, tant bactériens qu'eucaryotes, sont à l'origine, non seulement de réarrangements des chromosomes, mais également de la modulation de l'expression génique. L'élément mobile, grâce à ses propres signaux de transcription, va pouvoir contrôler l'expression des gènes à proximité de son site d'insertion. L'effet d'une insertion est souvent polaire (Shapiro, 1969). Ainsi, l'expression en aval de l'insertion peut être éteinte, ou au contraire activée comme c'est le cas pour Tn10 chez *Salmonella typhimurium*. Le premier effet résulte de la présence d'un terminateur fort de la transcription dans l'élément ; le deuxième effet est dû à la présence d'un promoteur dirigé vers l'extérieur de l'élément (Ciampi *et al.*, 1982; Bianco *et al.*, 1998).

De même chez les bactéries pathogènes, des points chauds mutationnels sont fréquemment observés en amont de gènes impliqués dans les relations avec l'hôte (adhérence, invasion). Ainsi, la variation du nombre de répétitions dans la région promotrice des gènes codant les antigènes de surface est à l'origine de la modulation de l'expression. Chez *Bordetella pertussis*, agent de la coqueluche, la variation de la taille d'une séquence composée de cytosines située en 5' de la phase codante entraîne une régulation différente de la transcription des gènes *fim* codant les antigènes de surface. Cette modulation est attribuée à la modification de la distance séparant le site de fixation d'un activateur de la transcription et la séquence -10 du promoteur (Willems *et al.*, 1990). Par ailleurs, des mécanismes d'inversion site-spécifiques de séquences d'ADN sont à l'origine de changements du cadre de transcription. Chez le bactériophage Mu, comme c'est le cas chez plusieurs autres bactériophages, des gènes sont alternativement exprimés par inversion et fusion à une séquence promotrice. Ce mécanisme entraîne la formation d'une diversité des protéines de queue de phage faisant varier le spectre d'hôtes (Dybvig, 1993).

Chez les bactéries, les éléments inversibles sont le plus souvent impliqués dans la régulation de l'expression de pili ou de flagelles. Les éléments identifiés sur des plasmides sont notamment impliqués dans la variation des pili ou d'autres facteurs jouant un rôle dans la dissémination des plasmides par conjugaison. Ainsi chez *Salmonella typhimurium*, l'expression des deux gènes *H1* et *H2* codant les flagellines, constituants protéiques majeurs des flagelles, dépend de l'orientation d'un segment d'ADN de 995 pb (Figure I-4.). Ce segment contient le promoteur d'un opéron constitué des gènes *H2* et *rb1*, ce dernier codant le répresseur de *H1*. Lorsque le promoteur est fonctionnel, *H2* et *rb1* sont exprimés, et conséquemment, *H1* est réprimé. Le mécanisme d'inversion est contrôlé par l'invertase Hin codée par l'élément de 995 pb et intervient à une fréquence de  $10^{-3}$  (Scott and Simon, 1982; Heichman and Johnson, 1990). Chez *E. coli*,

l'expression du gène *fimA*, codant la sous-unité majeure des protéines filamenteuses extracellulaires de type 1, est contrôlée par un mécanisme analogue (McClain *et al.*, 1991; Kulasekara and Blomfield, 1999). Ces protéines filamenteuses sont impliquées dans l'attachement des bactéries aux tissus hôtes.

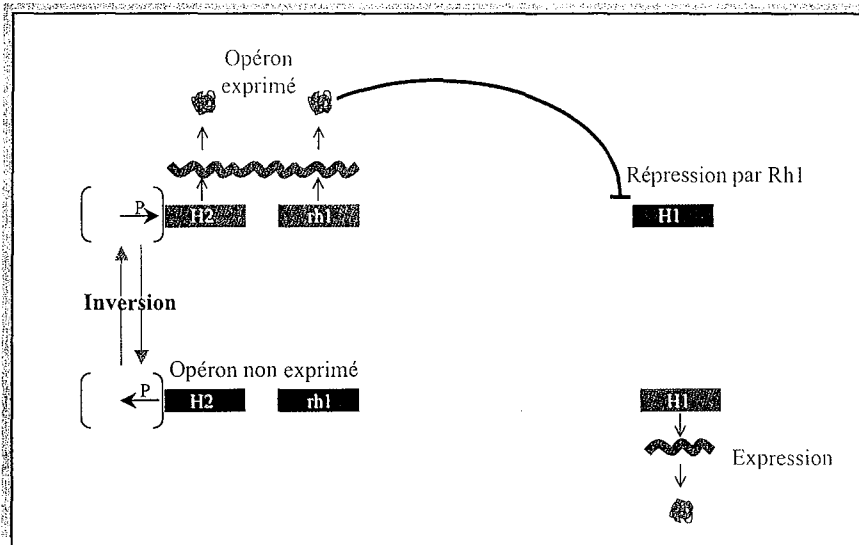


Figure I-4 : Illustration de réarrangements entraînant un changement des profils d'expression. Variation de phase chez *Salmonella*.

Les loci exprimés sont représentés en vert, les loci silencieux en bleu.  $\Rightarrow$  indique l'orientation du promoteur,  $\sim$  : ARN.  $\bullet$  : protéines.  $\{ \epsilon \}$  : élément inversible.

### A-I.3.2 - MODIFICATIONS DE LA TRADUCTION

Chez *Haemophilus influenzae*, agent de surinfection de la grippe et autres maladies respiratoires virales, les gènes impliqués dans la biosynthèse des lipopolysaccharides de surface contiennent un nombre variable (entre 16 et 19) de copies d'un motif tétramérique ( $5'$ CAAT $3'$ ) dans la région 5' traduite. La modification du nombre de copies par délétions ou duplications entraîne l'utilisation alternative de deux codons d'initiation de la traduction. Ainsi, les protéines codées diffèrent par leur région N-terminale (Szabo *et al.*, 1992; High *et al.*, 1993).

De plus, des variations de phénotype résultent de l'inversion d'un segment de gène à l'intérieur de séquences codantes. Ainsi, quatre segments d'ADN inversibles trouvés sur le plasmide conjugatif R64 d'*E. coli* sont responsables de la variabilité du pilus sexuel fin, impliqué dans la conjugaison en milieu liquide uniquement (Kim and Komano, 1997). En effet, les réarrangements résultant de l'inversion d'un ou plusieurs des quatre éléments, grâce à une recombinase site-spécifique (codée par le gène *rx1*), provoquent des variations de séquences dans la région 3' codante du gène *pilV* et donc des variations dans la partie C-terminale de la protéine PilV. Cette protéine est un composant mineur du pilus fin dont la synthèse est codée par le plasmide R64. La variabilité de la région C-terminale de cette protéine détermine la spécificité d'hôte.

## **A-I.4 $\omega$ MODULATION DE L'EXPRESSION PAR MODIFICATION DU DOSAGE GÉNIQUE**

La modification du dosage génique résulte de l'augmentation ou de la diminution du nombre de copies d'un gène par différents mécanismes : il peut s'agir de la répétition d'une séquence d'ADN portant le gène considéré ou encore de la modification de la position relative d'un gène par rapport à une origine de réplication. Un exemple classique de dosage de gène par surréplication est celui de l'amplification programmée des gènes du chorion de drosophile (Spradling and Mahowald, 1980). La surreprésentation des gènes codant les protéines du chorion, c'est-à-dire l'enveloppe de l'œuf, est assurée par réinitiations multiples de la réplication bidirectionnelle à partir d'une origine de réplication localisée à proximité de ces gènes. Cependant, aucun réarrangement du chromosome n'est associé à ce phénomène, le surdosage n'étant que transitoire.

De nombreux réarrangements chromosomiques, des bactéries aux eucaryotes supérieurs, entraînent une modification du dosage génique.

Le cas particulier des réarrangements de type duplication génique, pouvant entraîner des modulations de l'expression génique notamment par l'apparition de régions régulatrices distinctes pour les deux copies de gènes, sera détaillé dans le chapitre II.

### **A-I.4.1 - AMPLIFICATION DE GENES DE RESISTANCE AUX DROGUES DANS LES CELLULES EUCARYOTES**

L'amplification d'oncogènes est fréquemment rencontrée dans le développement des phases tardives de la cancérisation de nombreux organes chez l'homme (Debatisse and Tolédo, 1995). Des études quantitatives tendent à montrer que l'amplification est plus fréquente dans les lignées tumorales que dans les lignées saines. L'amplicon (qui correspond à l'unité amplifiable), dont la taille peut varier de 500 à 10 000 kb, porte des gènes codant des enzymes impliquées dans le métabolisme ou l'inactivation des drogues utilisées en chimiothérapie.

Les répétitions sont trouvées sous deux formes : des expansions chromosomiques initialement caractérisées par un spectre de bandes chromosomiques particulier, appelé HSR pour "Homogeneously Staining Region", et des mini-chromosomes acentriques circulaires, appelé DM pour chromosome "Double Minute". L'amplification peut être au locus chromosomique du gène amplifié ou être relocalisée probablement par l'intermédiaire des chromosomes DM.

La mutation du proto-oncogène p53 a été associée à la propension à l'amplification. La déficience de p53, donc de la signalisation de l'endommagement par l'amplification de la structure du

chromosome, permettrait aux cellules portant une amplification de survivre plutôt que d'être dirigées vers l'apoptose (Yin *et al.*, 1992).

### **A-I.4.2 - DOSAGE GENIQUE ET CONFIGURATION CHROMOSOMIQUE CHEZ LES BACTERIES**

Les exemples les mieux caractérisés concernent les bactéries pathogènes. En effet, bien que les phénomènes de réarrangements du génome puissent être observés chez toutes les bactéries, la pression de sélection exercée par l'hôte sur le polymorphisme naturel généré dans une population de bactéries permet de mettre clairement en évidence l'impact des réarrangements sur l'expression génique et donc la capacité d'invasion de l'hôte.

Chez *Salmonella typhi*, responsable de la fièvre typhoïde, un important polymorphisme de la structure génomique est observé. Il résulte d'événements de recombinaison homologue entre les sept opérons ribosomiques. Ces événements de recombinaison entraînent l'apparition de translocations et d'inversions de plusieurs centaines de kilobases (Liu and Sanderson, 1996; Ng *et al.*, 1999). Les réarrangements de type délétion et duplication, qui modifient fortement les distances entre origine et terminus de réplication, ne sont pas observés. Par contre, certaines inversions et translocations affectent la distance de certains gènes par rapport à l'origine de réplication. La proximité de l'origine de réplication entraîne une surreprésentation des gènes au cours de la phase exponentielle de croissance par réinitiations multiples à l'origine principale du chromosome *oriC*. Ainsi, le polymorphisme génomique mis en évidence chez les différents isolats de *S. typhi*, isolés au cours de l'étude de cas pathologiques, est interprété comme la sélection de variants génomiques dont le profil d'expression génique favorise l'invasion et la persistance chez l'hôte (Liu and Sanderson, 1996; NG *et al.*, 1999).

De la même façon, une diversité de la taille et du nombre de réplicons est observée chez les bactéries du genre *Brucella*. En effet, différentes configurations du génome sont trouvées chez des biovars de la même espèce *B. suis*, isolés d'hôtes animaux appartenant à des espèces distinctes. Le génome se présente sous la forme d'un chromosome circulaire unique de 3200 kb chez un isolat et sous la forme de plusieurs molécules circulaires chez les autres biovars. Cette redistribution de l'information génétique sur différents réplicons résulte d'événements de recombinaison homologue entre les trois loci *rrn* présents dans le génome. Les différentes configurations de l'information rendraient compte de la spécificité d'hôte (Jumas-Bilak *et al.*, 1998).

- CHAPITRE A-II -

LA DUPLICATION GENIQUE

## A-II.1 MÉCANISMES DE LA DUPLICATION DE GENES

Trois processus principaux peuvent être à l'origine de la duplication de gènes : la recombinaison homologue inégale, la polyploïdisation et la rétroposition.

### A-II.1.1 – LA RECOMBINAISON HOMOLOGUE INEGALE

Un échange non réciproque de matériel génétique entre chromosomes homologues peut amener à une duplication en tandem d'une région chromosomique. Chez les eucaryotes, ce phénomène peut se dérouler en méiose (Figure I-5.). La présence de séquences répétées de natures variées (éléments génétiques mobiles, gènes dupliqués ou ADN satellite) peut fournir le substrat de ces événements de recombinaison ectopique. Chez les procaryotes, le phénomène pourrait être observé entre deux chromatides sœurs issues de la réplication chromosomique ou dans des "compartiments cellulaires" comportant plusieurs copies du génome comme dans le cas des hyphes mycéliennes pluri-génomiques des *Streptomyces*.

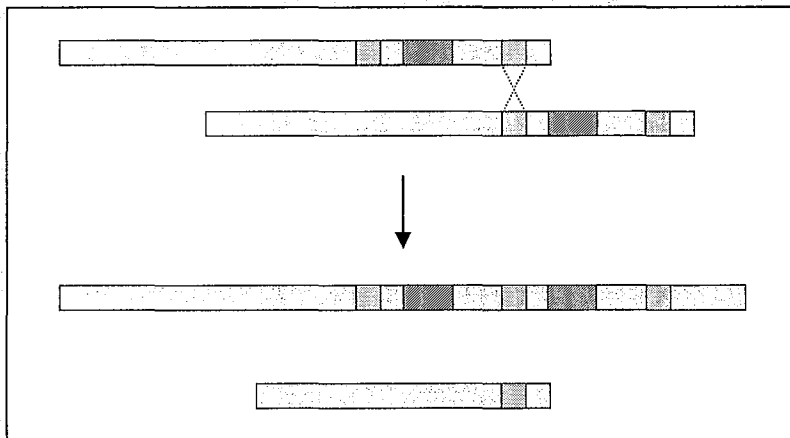


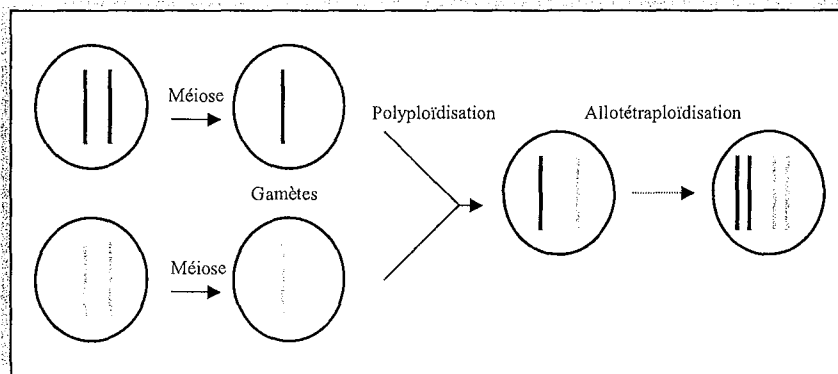
Figure I-5 : Mécanisme de duplication génique. Duplication de gènes par crossing-over non réciproque.

Recombinaison homologue entre séquences répétées (rectangle blanc pointillé) présentes le long d'un chromosome (rectangle gris). Après résolution, un des deux bras présente la duplication d'un gène (rectangle hachuré). Réciproquement, l'autre chromosome présente une délétion du gène considéré.

### A-II.1.2 – LA POLYPLOÏDISATION

Suite à la comparaison de génomes d'espèces appartenant à un même phylum, il est apparu que la duplication d'un génome entier par polyploïdisation (Figure I-6.) était un événement fréquent dans certaines lignées d'eucaryotes supérieurs comme les amphibiens, poissons (famille des Catostomidés) (Uyeno and Smith, 1972) ou végétaux (plus de 80% des angiospermes)

(Masterson, 1994), soit par autopolyploïdie (duplication du jeu de chromosomes de l'espèce) soit par allopolyploïdie (jeux chromosomiques multiples provenant d'espèces différentes). Les génomes modèles de *Saccharomyces cerevisiae* (Wolfe and Shields, 1997) et d'*Arabidopsis thaliana* (pour revue, Bancroft, 2001) dérivent vraisemblablement de la duplication du génome de l'espèce ancestrale. Cependant, seule une fraction limitée de gènes (entre 12,3% et 16% chez *S. cerevisiae* respectivement selon Seoighe et Wolfe (Seoighe and Wolfe, 1999; Wood *et al.*, 2002) aurait été retenue sous forme de duplicats dans les génomes contemporains suite aux nombreux réarrangements chromosomiques et notamment délétions ayant affecté le génome dupliqué.



**Figure I-6 : Mécanisme de duplication génique. Duplication génomique par polypléidisation.** Deux gamètes provenant d'espèces proches (cas d'allotétraploïdisation) fusionnent (les barres noires et grises représentent les jeux chromosomiques). Les chromosomes n'étant pas homologues mais homéologues (séquences nucléotidiques faiblement divergentes), le zygote n'est pas fertile par impossibilité de réaliser un appariement correct des chromosomes en méiose (prophase I). Un événement de polypléidisation permet la réalisation de bivalents en prophase I et rend ainsi possible le maintien de l'hybride.

### A-II.1.3 – LA RETROPOSITION

Un troisième mécanisme impliqué dans la duplication est la rétroposition.

Les transposons sont des éléments génétiques qui ont la faculté de se déplacer dans le génome. Ce déplacement peut impliquer directement une molécule d'ADN ou passer par l'intermédiaire d'une molécule d'ARN. Ce dernier mécanisme est appelé rétroposition. La rétroposition est donc définie comme un phénomène d'amplification de séquences d'ADN génomiques qui procède par transcription reverse des copies d'ARN suivie de l'insertion des ADNc dans différentes positions chromosomiques.

La rétroposition est un phénomène fondamental très conservé pour l'ensemble des eucaryotes. A l'origine, les rétroéléments répétés et dispersés dans le génome ont été classés en fonction de leur longueur : les éléments de quelques centaines de nucléotides ont été appelés SINE (Short

INterspersed Elements), ceux de plusieurs milliers de nucléotides LINE (Long INterspersed Elements).

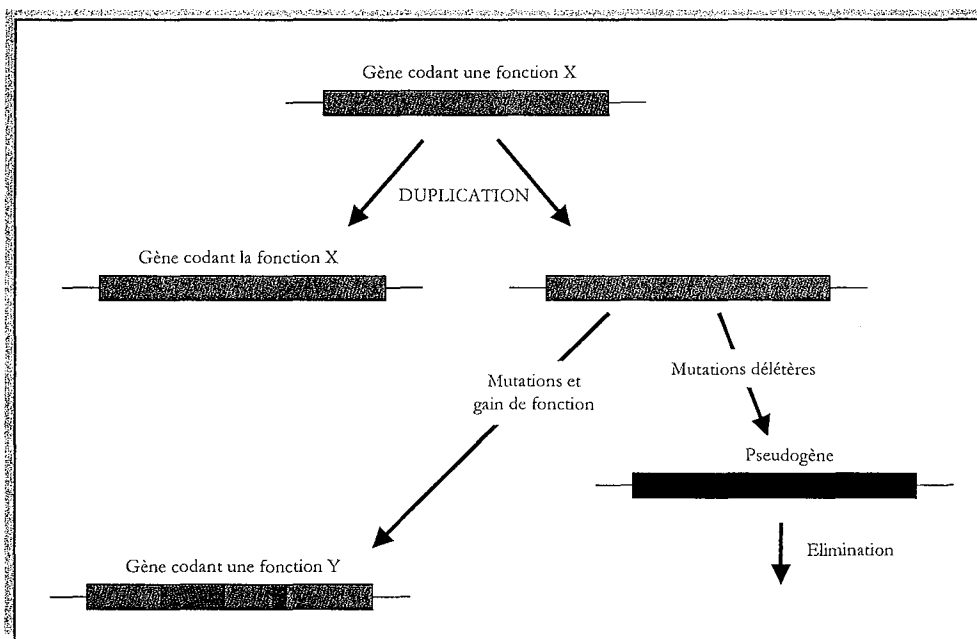
Les LINE sont des éléments autonomes qui codent pour les enzymes nécessaires à leur transposition, soit une transcriptase reverse et une, l'endonucléase. Les SINE sont des éléments non-codants qui pourraient parasiter le complexe de rétroposition LINE.

Bien caractérisée chez les eucaryotes supérieurs, cette activité n'a été identifiée que très rarement chez les bactéries (Inouye and Inouye, 1991). Cette activité a été décelée initialement chez *Myxococcus xanthus* (Inouye *et al.*, 1989) et *E. coli* (Lampson *et al.*, 1989) où la présence de séquences répétées a été interprétée comme pouvant résulter d'événements de rétroposition, c'est-à-dire de l'intégration dans le génome de fragments d'ADN complémentaire résultant d'une activité transcriptase inverse sur une molécule d'ARN.



## A-II.2 $\infty$ DUPLICATION ET EVOLUTION

Au travers des mécanismes d'évolution des gènes dupliqués, c'est la notion de "gène" qui est revisitée. En effet, le gène a été longtemps défini au niveau moléculaire comme une séquence d'ADN codante. Ainsi, dans le modèle classique de (Ohno, 1970) communément admis comme le paradigme de la théorie sur l'évolution des gènes, le seul mécanisme par lequel les membres d'une paire de gènes dupliqués peuvent échapper à la dégénérescence (formation d'un pseudogène) est la néofonctionnalisation. Par ce processus, une des copies acquiert une fonction conférant un avantage sélectif tandis que l'autre copie assure le maintien de la fonction originale (Figure I-7) (Ohno, 1970; Ohta, 1988; Walsh, 1995; Nowak *et al.*, 1997). Cependant, les gènes dupliqués peuvent être maintenus au sein d'un génome et présenter un autre devenir.



**Figure I-7 : Mécanismes d'évolution des gènes dupliqués selon (Ohno, 1970).**

Schéma illustrant le modèle classique qui conduit soit à la préservation d'une copie du gène dupliqué par une néofonctionnalisation (fonction Y), l'autre copie conservant la fonction initiale (fonction X), soit à la perte de la copie dupliquée par accumulation de mutations délétères (pseudogène).

### A-II.2.1 - LE MODELE DUPLICATION-COMPLEMENTATION-DEGENERESCENCE

Il est maintenant communément admis qu'un gène, s'il se caractérise par sa séquence (codante ou non), doit être considéré dans son environnement génomique et génotypique. En effet, un gène se définit également par son niveau d'expression dépendant de divers facteurs tels que son environnement génétique (local et chromosomique) et les gènes codant les effecteurs (protéines ou ARN) de son expression.

Ainsi, un gène est inclus dans un "réseau" complexe impliquant les différents éléments de régulation de l'expression génique. Le terme de réseau est une extension du concept de "cascades" de gènes selon lequel plusieurs gènes agissent les uns sur les autres de manière linéaire et hiérarchique. Le concept de réseau reconnaît l'existence de contrôles en retour ("feedback"), de redondances et de dialogues ("cross-talk") à l'intérieur et entre les cascades de gènes. Les gènes forment donc des unités de base qui composent un réseau de régulation (Shimeld, 1999). La transcription est régulée directement par l'intermédiaire de protéines, voire d'ARN, codés par d'autres gènes (dont les facteurs transcriptionnels) et indirectement par l'intermédiaire de la modulation de l'expression d'un gène du réseau (cas des molécules "signal", des récepteurs et des composants impliqués dans la transduction du signal ou la structuration de la chromatine).

De ce fait, deux copies exprimées dérivant d'un même gène ne doivent plus être considérées comme un seul et même gène. C'est sans doute une des raisons pour lesquelles le phénomène de duplication de gènes est considéré comme un processus primordial dans l'évolution au niveau moléculaire.

La plupart des génomes eucaryotes portent un nombre important de gènes dupliqués fonctionnels dont une grande majorité serait apparue il y a 10 à 100 millions d'années (Allendorf *et al.*, 1975; Cronn *et al.*, 1999). Suite à des événements massifs de duplication, de l'ordre de 20 à 50% des gènes dupliqués sont conservés. Cette conservation de la duplication durant des périodes de temps aussi longues suggère l'existence d'un mécanisme de sélection qui compenserait la production de pseudogènes (Li, 1985; Hughes and Hughes, 1993; Ramos-Onsins and Aguade, 1998).

Une hypothèse parcimonieuse concernant la conservation de gènes dupliqués à long terme est la sous-fonctionnalisation des deux membres de la paire. Par exemple, un gène exprimé à la fois dans deux tissus peut diverger vers deux copies, chacune étant exprimée uniquement dans l'un des tissus. Si ces deux sous-fonctions deviennent essentielles, un tel profil d'expression est retenu par la sélection naturelle. Les mutations gain-de-fonction, rares, n'ont plus besoin d'être invoquées puisque la sous-fonctionnalisation met en jeu des mutations dégénératives des séquences promotrices et stimulatrices (Force *et al.*, 1999).

Ce modèle prédit que des gènes dupliqués peuvent être maintenus pendant de très longues périodes avant d'acquiescer des sous-fonctions. Selon ce modèle, l'évolution de tels gènes a lieu en

deux phases (Figure I-8). Durant la phase I, qui se situe juste après l'événement de duplication, les loci dupliqués peuvent connaître trois destins alternatifs. Les deux premiers correspondent aux résultats prédits par le modèle classique. D'abord, une copie peut subir une mutation délétère dans sa région codante menant ultérieurement à sa perte (non-fonctionnalisation) ou dans ces éléments régulateurs (sous-fonctionnalisation). Un deuxième devenir concerne la néofonctionnalisation (modèle de Ohno). En principe, une néofonctionnalisation peut aussi être due à l'acquisition de régions régulatrices nouvelles sans pour autant altérer les fonctions existantes (Sidow, 1996). Le modèle DCD engloberait donc l'hypothèse de Ohno. Enfin, chaque copie peut connaître une perte ou une réduction de l'expression des sous-fonctions par des mutations dégénératives dans les séquences promotrices (substitutions, délétions, inversions, insertions). Dans ce cas, l'action combinée des deux loci dupliqués est nécessaire pour assurer la réalisation de la fonction ancestrale. La compensation qui s'effectue alors entre les deux gènes devrait préserver les deux copies partiellement dégénérées. Dans la deuxième phase, les gènes dupliqués peuvent diverger comme des gènes indépendants, notamment dans les régions régulatrices.

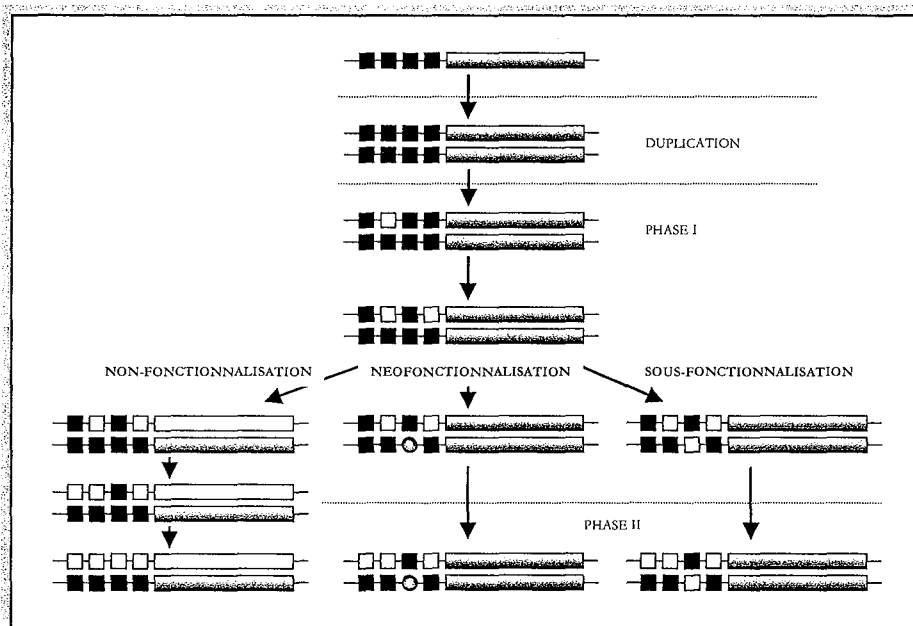
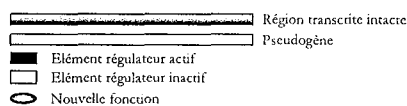


Figure I-8 : Application du modèle Duplication-Complémentation-Dégénérescence (d'après (Force *et al.*, 1999)). Les trois devenirs potentiels des deux copies d'un gène dupliqué comportant de multiples régions régulatrices. A gauche, les mutations dans les séquences promotrices fixes résultent de l'absence de fonction du pseudogène (abolition de la pression de sélection ou non-fonctionnalisation). A droite, la deuxième copie du gène acquiert une mutation délétère dans le troisième élément régulateur (sous-fonctionnalisation). Cette mutation préserve les deux copies d'une future perte de fonction car celles-ci sont maintenant essentielles pour l'expression génique complète. Au centre, une région régulatrice de la deuxième copie bénéficie d'une fonction adaptative qui préserve cette dernière (néofonctionnalisation).



## A-II.2.2 - DUPLICATION ET REGULATION DE L'EXPRESSION DES GENES DUPLIQUES

Cooke *et al.* (1997) ont observé que les gènes non essentiels, notamment ceux impliqués dans le métabolisme secondaire, seraient plus fréquemment dupliqués que les gènes de ménage ("housekeeping genes"). En effet, les gènes présentant une expression différentielle selon les tissus ou les stades de développement, et notamment les gènes régulateurs, peuvent acquérir de nouveaux domaines d'expression par mutation ou réarrangement de leur région promotrice ou activatrice (séquence "enhancer" chez les eucaryotes), et ce, sans variation de leur séquence protéique (Li and Noll, 1994; Hanks *et al.*, 1995).

De nombreux exemples de duplications géniques étayent ces hypothèses, et en particulier chez les eucaryotes. Chez les procaryotes, les exemples sont plus rares vraisemblablement parce que la taille des génomes étant plus contrainte, les événements de duplication génique pourraient y avoir été contre-sélectionnés au cours des processus évolutifs (Mushegian and Koonin, 1996).

### A-II.2.2.1 – DEVENIR DES GENES DUPLIQUES CHEZ LES EUCARYOTES

Depuis leur origine, voici quelques 600 à 800 millions d'années, les gènes de la globine chez l'homme donnent un panorama complet de l'évolution d'une super-famille de gènes (pour revue, voir (Goodman, 1981b; Goodman, 1981a)) avec des duplications de gènes, des divergences par des séries de mutations, l'acquisition de nouvelles fonctions et la perte de fonction (pseudogènes).

Cette superfamille comprend 2 familles et 1 gène, la famille des gènes de l' $\alpha$ -globine, la famille des gènes de la  $\beta$ -globine et la myoglobine, respectivement, situés sur trois chromosomes différents, Figure I-9).

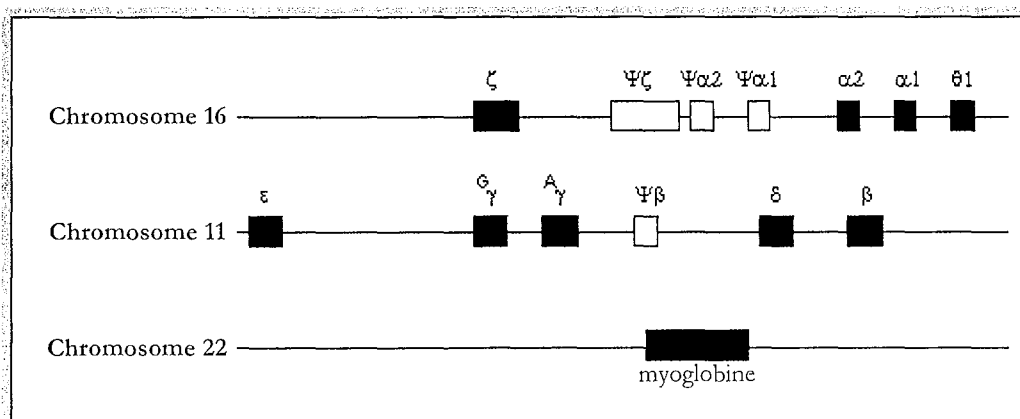


Figure I-9 : Les gènes de la globine chez l'homme.

L'ensemble de ces gènes présente des séquences nucléotidiques proches et une conservation de leur structure, témoignant qu'ils dérivent d'un gène ancestral par duplication (pour revue, voir (Goodman, 1981b; Goodman, 1981a)).

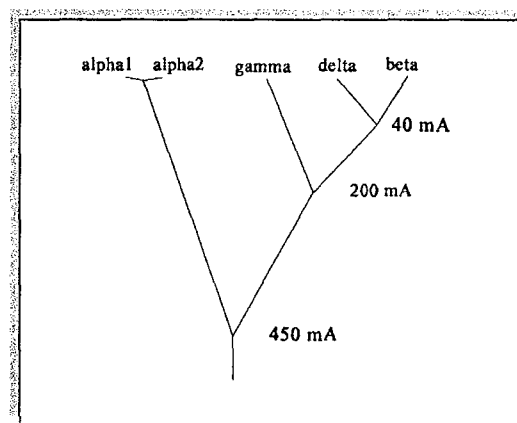
Les duplications peuvent être approximativement dater en comparant des espèces dont les dates de divergence sont supposées connues (Slightom *et al.*, 1980; Shen *et al.*, 1981; Powers and Smithies, 1986; Papadakis and Patrinos, 1999).

□ La divergence entre l' $\alpha$ - et la  $\beta$ -globine date d'environ 450 à 500 millions d'années, car on trouve ces 2 types de gènes chez tous les vertébrés sauf la lamproie. On pense donc que l'ancêtre commun des agnathes et des vertébrés devait posséder un seul gène de globine.

□ On retrouve la  $\beta$ -et la  $\delta$ -globine chez les primates uniquement, qui ont divergé entre 40 et 80 millions d'années.

□ On retrouve  $\alpha 1$  et  $\alpha 2$  distincts chez les seuls singes anthropoïdes qui ont divergé il y a moins de 20 millions d'années.

La Figure I-10 représente un modèle d'apparition des différents gènes d'apparition des différents gènes de la globine par duplications successives.



**Figure I-10 : Modèle d'apparition des gènes de la globine chez l'homme par duplications successives.**

Les datations des divers événements de duplications sont données en millions d'années (mA).

Ainsi, l'association de deux globines  $\beta$  et de deux globines  $\alpha$  conduit à l'hémoglobine adulte alors qu'une seule chaîne de myoglobine conduit à la protéine musculaire. En outre, les hémoglobines embryonnaires et foetales qui sont des tétramères comme l'hémoglobine adulte sont néanmoins formées de globines différentes. Ainsi, des gènes différents s'expriment à des moments différents du développement ou dans des tissus différents en assurant des fonctions pas nécessairement identiques puisque l'hémoglobine est une protéine de transport du dioxygène alors que la myoglobine en est une protéine de stockage.

De plus, à côté de ces gènes exprimés existent des pseudogènes qui ne conduisent pas à des protéines fonctionnelles mais dont la séquence se révèle également très proche de celles des autres globines.

### A-II.2.2.2 – DEVENIR DES GENES DUPLIQUES CHEZ LES BACTERIES

Chez les bactéries, les redondances de séquences sont particulièrement représentées par les ARN ribosomiques (loci *rnm*) et les séquences codant les ARN de transfert (Riley and Anilionis, 1978). Parmi les exemples de duplication de gènes codant pour des protéines chez les bactéries, deux devenirs de profils d'expression des duplicats peuvent être distingués. En effet, la duplication génique peut entraîner la surproduction d'une protéine ou la régulation différentielle l'expression d'une même protéine.

#### A-II.2.2.1 – Duplication de gènes aboutissant à la surproduction d'une protéine

Dans certains cas de duplication génique, l'analyse des profils d'expression de chaque copie de gènes n'a pu révéler de différence aussi bien en terme de régulation temporelle qu'en terme de régulation spatiale, suggérant ainsi que chaque gène contribuait à la production d'une grande quantité de protéine dans les mêmes conditions.

Dans le cas de la duplication des clusters de gènes codant la méthane monooxygénase chez *Methylococcus capsulatus* (99.996% d'identité), l'interruption d'un gène dans l'un ou l'autre des deux clusters entraîne une réduction du taux de croissance, indiquant alors que chaque copie est exprimée (Stolyar *et al.*, 1999).

L'inactivation par insertion des gènes *amoA* (99.879% d'identité) au sein des clusters de gènes dupliqués codant la monooxygénase à ammonium chez *Nitrosomonas europaea* a montré que les deux copies étaient exprimées (Hommes *et al.*, 1998). L'exemple des gènes *msa1* et *msa2*, codant l'antigène soluble majeur chez le pathogène *Renibacterium salmoninarum* rentre également dans ce cas de Figure où chaque copie présente une expression similaire (Rhodes *et al.*, 2002).

L'exemple le plus documenté concerne les gènes *tufA* et *tufB* qui contribuent, l'un comme l'autre, à la synthèse du facteur d'élongation EF-Tu, une protéine cytoplasmique très abondante chez *E. coli* (Furano, 1978). La majorité des bactéries Gram positives ne contiennent qu'une seule copie de ce gène *tuf*, alors que les bactéries Gram négatives, telles qu'*E. coli*, présentent deux copies de ce gène avec une très forte identité nucléotidique (par exemple 99% chez *S. typhimurium*; (Furano, 1978), et *E. coli*; (Jaskunas *et al.*, 1975)). Une exception est cependant représentée par le cas de *Streptomyces ramocissimus*, chez qui trois copies de ce gène *tuf* ont pu être identifiées (Vijgenboom *et al.*, 1994).

**A-II.2.2.2.2 – Duplication de gènes aboutissant à l'expression différentielle d'une protéine**

Parmi les exemples de d'expression différentielle de gènes dupliqués chez les bactéries, nous pouvons citer le cas de la phosphoribulokinase chez *Alcaligenes eutrophus* (Klintworth *et al.*, 1985), ou encore de la ribulose-1,5-bisphosphate carboxylase chez *Rhodobacter sphaeroides* 2.4.1 (Hallenbeck *et al.*, 1990).

Ce paragraphe détaillera quelques cas connus de telles régulations différentielles de gènes dupliqués chez les bactéries modèles *E. coli* et *S. coelicolor*.

**A-II.2.2.2.2.1 – Les gènes *lysS* et *lysU* chez *E. coli***

Chez *E. coli* K12, deux gènes dupliqués codant des lysyl-tRNA synthétases ont été identifiés (Hirshfield *et al.*, 1981). L'un de ces deux gènes, nommé *lysS*, est exprimé constitutivement tandis que le second, *lysU*, est induit dans différentes conditions environnementales, incluant un stress thermique ou l'addition de leucine dans le milieu. Un mutant interrompu pour le gène *lysU* ne présente pas de phénotype particulier, indiquant que la copie *lysU* n'est pas essentielle à la croissance (Clark and Neidhardt, 1990). Par ailleurs, l'interruption du gène *lysS* entraîne l'altération de la croissance à des faibles températures. Cette sensibilité aux basses températures peut être complémentées par l'introduction d'un plasmide multicopie portant le gène *lysU* (Leveque *et al.*, 1991; Kawakami *et al.*, 1992). Ainsi, le gène *lysS* apparaît comme « dispensable » dans des conditions de croissance à hautes températures, permettant l'induction de l'expression du gène *lysU*. En revanche, à des faibles températures, le gène *lysS* devient « essentiel », de par la non expression du gène *lysU*. L'activité lysyl-tRNA synthétase est donc elle-même essentielle à la viabilité des cellules, et *lysS* peut être fonctionnellement remplacée par *lysU*.

La régulation différentielle de ces deux gènes pourrait donc être à l'origine du maintien de leur duplication (Kawakami *et al.*, 1992).

**A-II.2.2.2.2.2 – Les gènes *csp* chez *E. coli***

Chez *E. coli*, les gènes *cspA* à *cspI* forment une famille de 9 gènes dont les fonctions ont divergé et ont ainsi permis une adaptation aux variations des conditions environnementales (Yamanaka *et al.*, 1998). Le premier gène caractérisé, *cspA*, code la protéine chaperonne majeure de la réponse au choc froid. La protéine CspA semble faciliter la traduction en empêchant la formation de structures secondaires au niveau de l'ARN. CspB et CspG sont également synthétisées en réponse aux basses températures mais leur rôle reste inconnu. CspD est impliquée dans le stress de type nutritionnel tandis que CspC et CspE auraient une fonction dans la régulation transcriptionnelle ou la condensation du chromosome. Aucune fonction n'a encore été attribuée à CspF, CspH et CspI. L'analyse des séquences géniques suggère fortement que ces copies de gènes dérivent d'un phénomène de duplications successives. Ces neuf gènes n'auraient pas encore assez divergé pour que leur origine commune ne soit plus décelable.

### A-II.2.2.2.3 – Régulation différentielle de gènes dupliqués chez les Streptomyces

Chez *S. coelicolor* A3(2), de nombreux exemples de gènes dupliqués présentant une expression différentielle en fonction des tissus ou des conditions environnementales sont disponibles. Citons par exemple le cas de l'opéron glycogène dont une copie (*glgBI*) participe au métabolisme du glycogène en phase de croissance végétative, et l'autre (*glgBII*) est impliquée dans l'accumulation de glycogène dans les spores et est exprimée en phase tardive de la différenciation cellulaire (Bruton *et al.*, 1995). Egalement, deux gènes codant deux superoxyde dismutases de même type (FeSOD pour "Fer-containing SuperOxide Dismutase") présentent un niveau de régulation par le nickel différent (Chung *et al.*, 1999).

## **A-II.2.3 - RELATION ENTRE LINEARITE CHROMOSOMIQUE, PLASTICITE GENOMIQUE ET DUPLICATION GENIQUE**

Chez les eucaryotes comme chez les procaryotes, certaines régions chromosomiques semblent être le siège de nombreux réarrangements et constitueraient de véritables zones d'essai permettant, d'une part, à une duplication d'apparaître et de persister du fait du relâchement des contraintes sélectives dans ces parties du génome, et d'autre part, d'éprouver le caractère avantageux de nouvelles fonctions apparues spontanément grâce à la plasticité génomique.

### A-II.2.3.1 – POLYMORPHISME DES REGIONS SUBTELOMERIQUES

#### A-II.2.3.1.1. - Variation des régions subtélomériques eucaryotes

Sur les chromosomes linéaires eucaryotes, les régions permissives aux réarrangements sont localisées à proximité des centromères ou plus généralement dans les régions subtélomériques. Ces régions subtélomériques sont polymorphes et composées dans la majorité des chromosomes eucaryotes d'une combinaison de gènes et de séquences variables répétées en tandem ou dispersées (pour revue voir (Pryde *et al.*, 1997).

Ainsi, l'analyse de la séquence génomique de *S. pombe* (Wood *et al.*, 2002) révèle l'existence de régions de syntenie intragénomique, c'est-à-dire de régions dupliquées d'organisation génétique conservée. Ces régions d'environ 50 kb sont situées dans les régions subtélomériques des chromosomes I et II. Chez *S. cerevisiae*, 7 gènes parmi 16 possédant une identité en nucléotides de 100% sont également situés dans les régions subtélomériques. La plupart de ces gènes coderaient des protéines de surface non essentielles spécifiques de chaque espèce. Chez les eucaryotes pathogènes et parasites, comme par exemple chez l'agent causal de la malaria, *Plasmodium*



*falciparum*, les régions subtélomériques contiennent la majorité des gènes codant pour les antigènes de surface (Bowman *et al.*, 1999). Chez ces organismes, les réarrangements des régions subtélomériques conduisent à la variation de l'expression d'antigènes de surface et permettent ainsi au parasite d'échapper à la réponse immune de l'hôte (Rubio *et al.*, 1996; Wada and Nakamura, 1996; Freitas-Junior *et al.*, 2000).

De même, chez l'homme, plusieurs exemplaires d'un même gène dont un semble fonctionnel, appartenant à la famille des récepteurs olfactifs, sont localisés dans les régions subtélomériques (Linardopoulou *et al.*, 2001). De plus, de nombreux loci polymorphes dans la population humaine sont localisés dans les régions subtélomériques (par exemple, le locus codant l' $\alpha$ -globine, (Wilkie *et al.*, 1991).

#### **A-II.2.3.1.2. – Régions permissives aux réarrangements chez les bactéries**

##### *A-II.2.3.1.2.1 – Réarrangements dans les chromosomes circulaires bactériens*

Chez certaines espèces bactériennes, la comparaison de l'organisation génomique de différentes souches montre que toutes les régions du chromosome ne présentent pas la même permissivité aux réarrangements chromosomiques. L'accumulation de remaniements dans une partie restreinte du chromosome entraîne l'existence d'une région variable et une région mieux conservée.

Ainsi, le chromosome de *Pseudomonas putida* (2,8 Mb) est environ deux fois plus petit que celui de *P. aeruginosa*. Cette différence serait essentiellement due à l'absence de certaines séquences dans la région de terminaison de la réplication du chromosome de *P. putida* (Krawiec and Riley, 1990; Kolsto, 1997).

La comparaison de la carte génétique et physique du chromosome des espèces *Neisseria gonorrhoeae* et *N. meningitidis*, deux espèces pathogènes pour l'homme, montre l'existence de réarrangements complexes, de type translocation et inversion, localisés essentiellement dans la région de terminaison de la réplication (Dempsey *et al.*, 1995).

Chez *Bacillus cereus*, une partie du chromosome, localisée dans la région de terminaison de la réplication, présente un polymorphisme entre souches d'une même espèce (Carlson and Kolsto, 1994).

Ainsi, chez les bactéries, les régions de terminaison de la réplication sont le siège de nombreux réarrangements et constitueraient un substrat idéal pour la constitution de nouvelles fonctions.

A-II.2.3.1.2.1 – Polymorphisme des régions subtélomériques des chromosomes linéaires bactériens

De façon analogue aux eucaryotes pathogènes, chez *B. hermsii*, un gène codant des protéines de surfaces est adjacent aux télomères. Alors que plusieurs allèles de ces gènes sont retrouvés sur des plasmides, seul le site chromosomique permet son expression (Kitten and Barbour, 1990; Restrepo *et al.*, 1992; Kitten *et al.*, 1993).

Chez *Borrelia burgdorferi*, de fortes similarités de séquences existent sur plusieurs kilobases au niveau des régions subtélomériques entre les 13 réplicons linéaires composant le génome (12 plasmides linéaires et le chromosome (Casjens *et al.*, 2000). De plus, les séquences du subtélomère chromosomique droit de *B. burgdorferi* sont souvent retrouvées au niveau des régions subtélomériques des plasmides de cette espèce alors que l'ensemble des autres séquences chromosomiques est spécifique du chromosome (Casjens *et al.*, 1997).

Les régions terminales du chromosome linéaire des *Streptomyces* correspondent à des termini naturels de la réplication et sont caractérisées par une très forte instabilité structurale (Leblond and Decaris, 1999) voir chapitre A-III).

De plus, les répétitions terminales inversées chromosomiques des *Streptomyces* présentent une très grande variabilité au niveau de leur taille, comprise entre 174 pb pour *S. avemitilis* (Omura *et al.*, 2001) à 550 kb pour *S. griseus* (Lezhava *et al.*, 1995b). Chez *S. coelicolor* A3(2), ces répétitions mesurent 21653 pb (Bentley *et al.*, 2002) et chez *S. ambofaciens*, leur taille a été estimée à environ 210 kb (Leblond *et al.*, 1996). Les répétitions terminales inversées plasmidiques présentent la même variabilité. Ainsi la taille des répétitions terminales inversées des plasmides des *Streptomyces* varie de 44 pb (plasmide SLP2 de *S. lividans*, (Chen *et al.*, 1993) à 180 kb (plasmide pZG101 de *S. rimosus* (Pandza *et al.*, 1998).

Peu d'homologie existe, en général, entre les séquences des répétitions terminales inversées, que ce soit pour les chromosomes et les plasmides. Par exemple, aucune homologie ne peut être détectée par hybridation entre les répétitions terminales inversées de *S. coelicolor* et *S. lividans* (Huang *et al.*, 1998), bien que ces deux espèces soient phylogénétiquement proches (Takeuchi *et al.*, 1996). La variabilité des séquences subtélomériques existe également au niveau intraspécifique. En effet, la cartographie comparée des TIR de *S. ambofaciens* DSM40697 et ATCC23877 montre l'existence de séquences spécifiques de chaque isolat dans les derniers kilobases des chromosomes (voir Discussion), suggérant l'existence d'échanges d'information génétique (transfert horizontal) dans le biotope naturel des *Streptomyces* qu'est le sol.

L'analyse de la séquence complète du génome de *S. coelicolor* A3(2) (Bentley *et al.*, 2002) a permis de révéler une véritable compartimentation de l'information génétique, malgré une densité en gènes similaire tout au long du chromosome, y compris dans la région instable touchée par les réarrangements.

En effet, une forte concentration des gènes dits « essentiels » (i.e. gènes impliqués dans les phénomènes nécessaires à la croissance végétative : gènes codant des fonctions impliquées dans

la division cellulaire, la réplication de l'ADN, la transcription, la traduction et la biosynthèse des acides aminés) a été détectée dans une région d'environ 4,9 Mb constituant le « core » chromosomique (Bentley *et al.*, 2002). Cette région « core » contient l'origine de réplication chromosomique et correspond à la partie centrale du chromosome (Figure I-11).

Les « bras chromosomiques » sont définis comme les régions terminales gauche (1,5 Mb) et droite (2,3 Mb), dans lesquelles les gènes « essentiels » sont rares mais où sont concentrés les gènes dits de "contingence" (Figure I-11). Les gènes de « contingence », pour ceux dont la fonction a été définie, sont impliqués dans la différenciation cellulaire, dans l'expression du métabolisme secondaire et l'adaptation aux changements environnementaux (biotiques ou abiotiques). De plus, les régions de « contingence » sont particulièrement riches en ORF codant des transposases putatives. En effet, environ 40% de la totalité des ORF codant des transposases putatives sont situées dans les 200 derniers kilobases constituant les extrémités chromosomiques, soit une densité 17 fois supérieure à celle du reste du génome. Ceci suggère que les régions terminales du chromosome linéaire soient plus permissives que les autres à l'intégration des transposons et IS (Bentley *et al.*, 2002).

Les rares gènes "essentiels" identifiés dans les deux bras chromosomiques correspondent le plus fréquemment à un duplicat d'une ORF localisée dans la région « core » (voir ci après, paragraphe II.2.3.2).

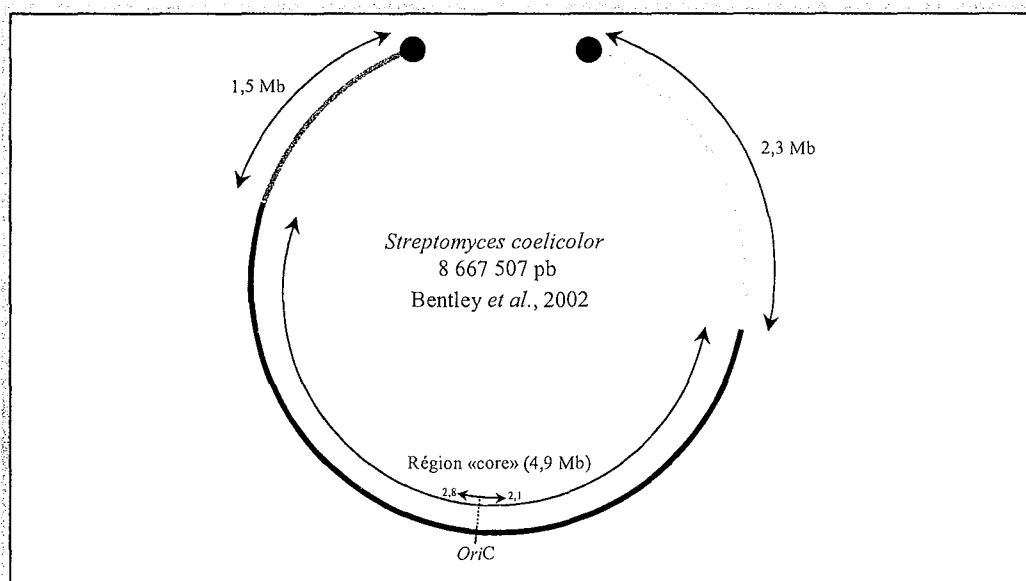


Figure I-11 : Compartimentation du chromosome de *S. coelicolor* A3(2).

Selon (Bentley *et al.*, 2002).

La région "core" présente une forte densité en gènes "essentiels" et englobe l'origine de réplication chromosomique. Les bras chromosomiques gauche (gris foncé) et droit (gris clair) correspondent aux régions terminales du chromosome. Les cercles noirs représentent les protéines terminales (Tpg). La taille des régions est indiquée en mégabases (Mb).

Enfin, les comparaisons préliminaires des séquences des répétitions terminales inversées de *S. ambofaciens* ATCC23877 (projet Génoscope-IGM Orsay-LGM Nancy) avec celles de *S. coelicolor* montrent que la majorité des ORF détectées chez *S. ambofaciens* n'ont pas d'homologue fort dans le génome de *S. coelicolor* (voir Discussion).

Ainsi, l'organisation génétique pourrait être fortement conservée dans les régions internes du chromosome linéaire des *Streptomyces* (région « core »). En revanche, les régions terminales seraient permissives aux réarrangements.

### A-II.2.3.2 – LINEARITE CHROMOSOMIQUE ET TAUX DE REDONDANCE GENOMIQUE

#### A-II.2.3.2.1 – Taux de redondance des génomes séquencés

L'analyse de la séquence génomique par le programme BLAST-CLUST (Altschul *et al.*, 1997), réalisée au cours de ce travail, indique que 8,8% des ORF chez *S. coelicolor* A3(2) et 7,9% des ORF chez *S. avermitilis* seraient issus d'événements de duplication génique (respectivement 7717 et 7460 ORF considérées qui excluent les transposases d'éléments génétiques mobiles).

Le seuil de 50% d'identité entre les protéines déduites de séquences génomiques est appliqué lors de cette étude. Ces duplications peuvent être simples, créant des duplicats ou multiples, générant des familles des gènes.

Ces valeurs placent les deux bactéries du genre *Streptomyces* immédiatement derrière *S. cerevisiae* (12,3%), et *Borrelia burgdorferi* (26,4%) et devant la levure *Schizosaccharomyces pombe* (7,4%), (Wood *et al.*, 2002). Alors que *S. cerevisiae* représente le modèle d'étude d'évolution par duplication génomique chez les eucaryotes (Wolfe and Shields, 1997), les *Streptomyces* correspondent aux génomes bactériens de grande taille, actuellement séquencés, les plus redondants.

Il est également intéressant de souligner que cette analyse semble associer étroitement la linéarité chromosomique avec le niveau de redondance génique (Tableau I-1). Ainsi, la bactérie *B. burgdorferi*, agent de la maladie de Lyme (Ferdows and Barbour, 1989) présente un chromosome linéaire (1 Mb) et une redondance génomique très forte principalement révélée par la présence de familles de gènes représentant au total environ 26% des ORF du génome.

Espèce	Phylum	Taille		Nombre d'ORF*	Duplicats (%)	Familles (%)	Gènes en copie unique (%)	Source	Réf
		du génome (kb)							
<i>Schizosaccharomyces pombe</i>	Eucaryote / champignon	13800	4876	5,1	7,4	92,6	S	(1)	
<i>Saccharomyces cerevisiae</i>	Eucaryote / champignon	12156	5777	8,9	12,3	87,7	S	(2)	
<i>Streptomyces coelicolor</i> A3(2)	Gram positive Haut G/C	8667	7717	6,3	8,8	91,2	S	(3)	
<i>Streptomyces avermitilis</i>	Gram positive Haut G/C	9026	7460	5,8	7,9	2,19	A	(4)	
<i>Escherichia coli</i> K12	Protéobactérie Subdivision $\gamma$	4600	4233	5,4	6,7	93,3	T	(5)	
<i>Mycobacterium tuberculosis</i>	Gram positive Haut G/C	4400	3824	3,9	6,6	93,4	T	(6)	
<i>Bacillus subtilis</i>	Gram positive Bas G/C	4200	4095	5,8	7,1	92,9	T	(7)	
<i>Agrobacterium tumefaciens</i> C58	Protéobactérie Subdivision $\alpha$	2800	5271	2,8	4,9	95,1	T	(8)	
<i>Staphylococcus aureus</i>	Gram positive Bas G/C	2800	2690	4,9	7	93	T	(9)	
<i>Lactococcus lactis</i> IL403	Gram positive Bas G/C	2300	2216	5,5	7,6	92,4	T	(10)	
<i>Borrelia burgdorferi</i>	Spirochètes	910	1635	3,2	26,4	73,6	T	(11)	
<i>Mycoplasma genitalium</i>	Gram positive Bas G/C	580	480	0	0	100	T	(12)	

Tableau I-1 : Estimation du niveau de redondance des génomes totalement séquencés.

Le programme BLAST-CLUST est une communication personnelle de Yuri Wolf. Les séquences génomiques ainsi que les fichiers d'ORF déduites sont accessibles sur les sites WEB de « l'Institute for Genomic Research » (T) (TIGR, <http://www.tigr.org/>), du « Sanger Institute » (S) (<http://www.sanger.ac.uk/>) ou sur le site correspondant au projet de séquençage du génome de *S. avermitilis* (A) (<http://avermitilis.ls.kitasato-u.ac.jp/>) La duplication génique est déduite de l'observation d'une identité minimale de 50% entre deux protéines (duplicats) ou plus (famille de gènes) codées par un génome donné.

\* : les ORF identifiées comme étant des transposases ont été retirées des échantillons.

(1) Wood *et al.*, 2002 ; (2) Goffeau *et al.*, 1997 ; (3) Bentley *et al.*, 2002 ; (4) Ikeda *et al.*, 2003 ; (5) Blattner *et al.*, 1997 ; (6) Cole *et al.*, 1998 ; (7) Kunst *et al.*, 1997 ; (8) Goodner *et al.*, 2001 ; (9) Kuroda *et al.*, 2001 ; (10) Bolotin *et al.*, 1999 ; (11) Fraser *et al.*, 1997 ; (12) Fraser *et al.*, 1995.

### A-II.2.3.2.2 – Localisation des duplicats ou familles de gènes

De plus, il est intéressant de noter la localisation particulière des duplicats ou des membres de famille de gènes sur les chromosomes de *E. coli* et de *S. coelicolor* A3(2). Ainsi, dans un grand nombre des cas, une copie est localisée dans la région centrale du chromosome alors que l'autre copie est localisée dans les régions du terminus.

Par exemple, pour les deux opérons homologues participant à la synthèse du glycogène chez *S. coelicolor* A3(2) (voir paragraphe A-II.2.2.2.2.3, (Bruton *et al.*, 1995)) *glgBI*, participant à la croissance végétative, est localisé dans la région « core » alors que *glgBII*, impliqué dans le processus de maturation des spores, est localisé dans un « bras chromosomique ». Le premier assure une fonction « essentielle » alors que le second peut être considéré comme « contingent ». Une analyse analogue peut être illustrée avec les gènes codant les superoxydes dismutases (SOD) (Chung *et al.*, 1999).

Chez *E. coli*, dans le cas des gènes *asp* mentionnés précédemment (voir paragraphe A-II.2.2.2.1.3), le gène *aspA* se situe à proximité de l'origine de réplication alors que les huit autres membres sont localisés de part et d'autre du site de terminaison de la réplication (Yamanaka *et al.*, 1998).

De plus, les 486 ORF dupliquées du génome de *S. coelicolor* selon l'analyse par le programme BLASTCLUST (voir Tableau I-1, 6,3% de 7717 ORF) ont été localisées sur la carte chromosomique de *S. coelicolor*. Dans la majorité des cas, lorsque une des deux copies est localisée dans un « bras chromosomique », la seconde se retrouve dans la région « core » du chromosome (Figure I-12). Il est à noter la surreprésentation particulière de ces duplicats sur le bras chromosomique droit comparativement au bras chromosomique gauche (Figure I-12).

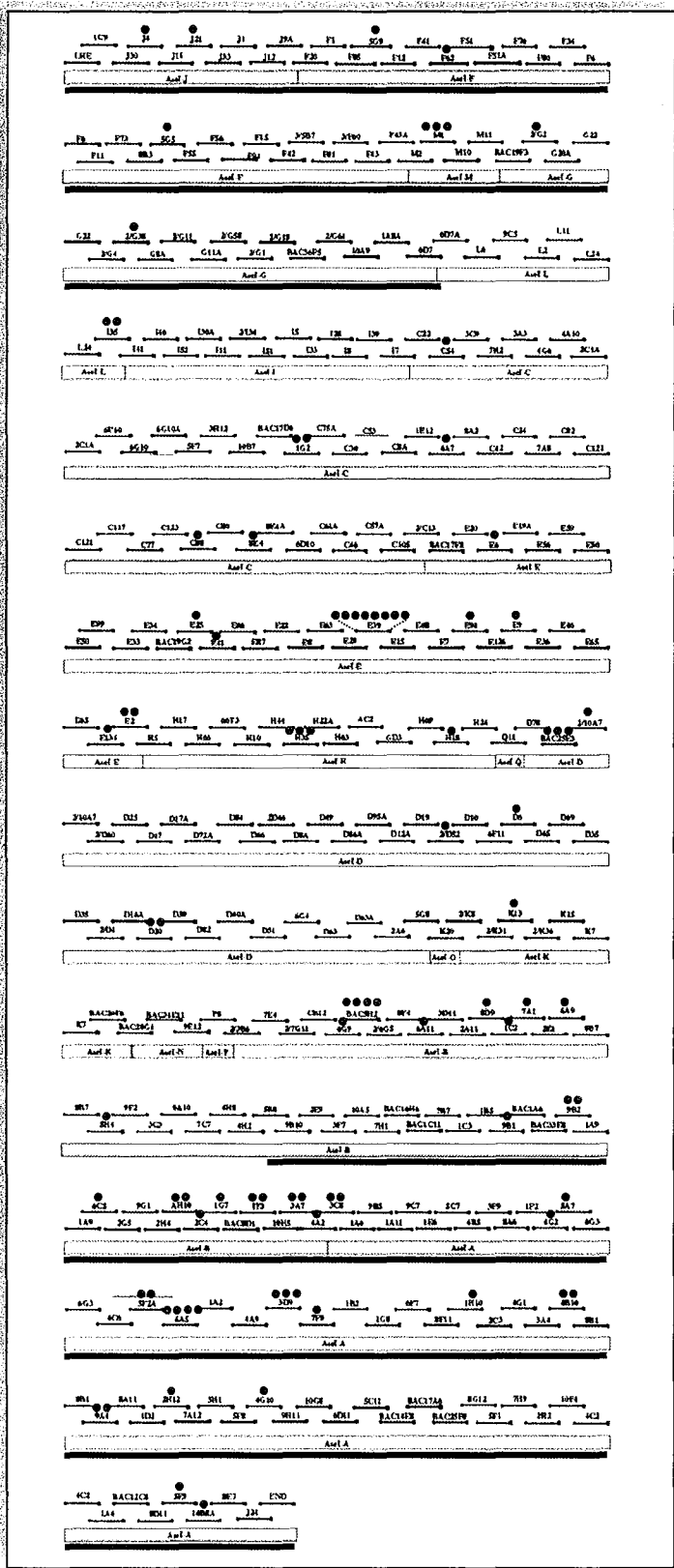


Figure I-12 : Localisation des gènes dupliqués sur le chromosome de *S. coelicolor* A3(2). Les bras chromosomiques sont représentés en bleu, les duplicats par des points rouges.

**A-II.2.3.2.3 – Fonctions redondantes**

Les fonctions des protéines putatives codées par les gènes dupliqués chez *S. coelicolor* (8,8% de 7717 ORF) sont présentées dans le Tableau I-2. Ainsi, le pourcentage de chaque classe de protéines apparaissant comme redondante dans cette analyse a été déterminé (Tableau I-2).

<b>Fonction</b>	<b>%</b>
<u>Fonction inconnue</u>	<u>21,63</u>
<u>Processus cellulaires</u>	<u>9,62</u>
Protection	1,92
Protéines de transport	6,73
<u>Métabolisme des macromolécules</u>	<u>2,88</u>
Dégradation des macromolécules	2,40
Biosynthèse, modification des macromolécules	0,48
<u>Métabolisme des petites molécules</u>	<u>23,56</u>
Biosynthèse des acides aminés	1,44
Biosynthèse des co-facteurs	0,96
Métabolisme central	3,85
Dégradation des petites molécules	6,25
Métabolisme énergétique, du carbone	5,77
Biosynthèse des acides gras	
Biosynthèse des nucléotides	
Métabolisme secondaire	5,29
<u>Enveloppe cellulaire</u>	<u>10,58</u>
Lipoprotéines (périplasmiques, exportées)	9,62
Constituants du ribosome	0,96
<u>Extrachromosomique</u>	<u>7,21</u>
Fonctions relatives aux plasmides	7,21
<u>Régulation</u>	<u>16,35</u>
Systèmes de régulation à deux composants	2,88
Facteurs sigma	2,88
Familles de régulateurs	8,17
<u>Autres</u>	<u>2,40</u>
<u>Non classés</u>	<u>8,17</u>

Tableau I-2: Pourcentages des familles de protéines redondantes dans le génome de *S. coelicolor*.

Selon la classification des protéines de l'Institut Sanger.



Il est à noter que sur les 160 classes de protéines répertoriées, seule une vingtaine apparaît comme redondante selon cette analyse. La duplication génique ne semble donc affecter n'importe quel type de fonction. Les systèmes de régulation notamment, semblent être issus d'événements de duplications géniques puisque cette famille de protéines représente environ 16% des fonctions redondantes chez *S. coelicolor*.

Lors de l'analyse du génome de *S. avermitilis* (Ikeda *et al.*, 2003), les familles de gènes ont été caractérisées à l'aide du même programme BLAST-CLUST. Les paramètres appliqués par les auteurs sont cependant différents de ceux utilisés dans ce travail, avec une identité minimale de 30% (contre 50% ici) sur un minimum de recouvrement de 80% (contre 90% ici). Ainsi, 35% des ORF codées par le génome de *S. avermitilis* peuvent être regroupées dans 721 familles de gènes paralogues, avec entre 2 et 91 gènes par famille (Ikeda *et al.*, 2003).

Les trois plus grandes familles de gènes sont représentées par les ABC transporteurs, les systèmes de régulation transcriptionnelle à deux composants et les facteurs sigma.

En effet, 60 facteurs  $\sigma$  putatifs ont pu être décelés dans le génome de *S. avermitilis*, 47 d'entre eux appartenant à la famille des ECF (voir paragraphe B-II.1, ECF pour "extracytoplasmic function sigma factors" ; (Lonetto *et al.*, 1994; Missiakas and Raina, 1998). De même, 65 gènes putatifs codant des facteurs  $\sigma$  ont été décelés au cours du séquençage systématique du génome de *S. coelicolor* A3(2), 45 d'entre eux étant de type ECF (Bentley *et al.*, 2002).

Bien que le rôle de chacun de ces facteurs  $\sigma$  reste à définir, l'analyse de leur séquence révèle l'existence de familles de gènes issus d'événements successifs de duplication (Figure I-13). Une famille de 9 gènes codant des facteurs  $\sigma$  homologues du facteur  $\sigma$  de réponse générale au stress  $\sigma^B$  de *Bacillus subtilis* fera l'objet de nos recherches (voir Résultats).

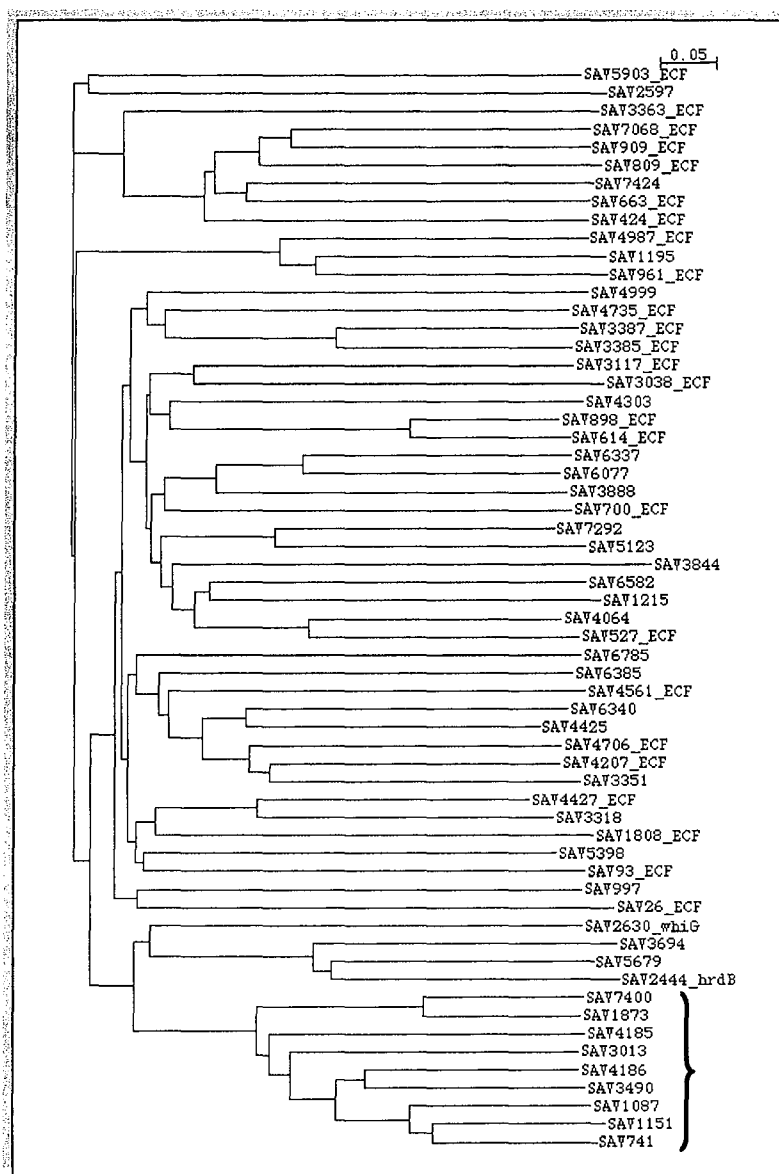


Figure I-13 : Arbre phylogénétique des 60 gènes putatifs codant des facteurs  $\sigma$  chez *S. avermitilis*.

Selon (Ikeda *et al.*, 2003).

Les neuf gènes repérés par l'accolade rouge codent des facteurs  $\sigma$  homologues du facteur  $\sigma^B$  de *Bacillus subtilis*.

Cette pléthore de facteurs  $\sigma$  fait de *Streptomyces* le taxon bactérien le plus riche en facteurs transcriptionnels de ce type (Tableau I-3). Cette caractéristique est à corrélérer à la grande complexité du processus de différenciation (biochimique et morphologique), ainsi qu'aux nombreux régulons de réponse aux stress notamment. Le nombre des facteurs  $\sigma$  chez les bactéries, y compris dans les genres phylogénétiquement apparentés au genre *Streptomyces* (par exemple *Mycobacterium*), suggère fortement que l'émergence de la complexité de la régulation transcriptionnelle par les facteurs  $\sigma$  alternatifs serait d'acquisition évolutive récente.

Espèce	Phylum	Nombre d'ORF*	Nombre de facteurs $\sigma$ putatifs	Ratio nombre de $\sigma$ / ORF (X 1000)	Réf
<i>Streptomyces coelicolor</i> A3(2)	Gram positive Haut G/C	7717	65	8,55	(1)
<i>Streptomyces avermitilis</i>	Gram positive Haut G/C	7460	60	8	(2)
<i>Escherichia coli</i> K12	Protéobactérie Subdivision $\gamma$	4233	7	1,5	(3)
<i>Mycobacterium tuberculosis</i>	Gram positive Haut G/C	3824	13	2,9	(4)
<i>Bacillus subtilis</i>	Gram positive Bas G/C	4095	16	3,8	(5)
<i>Agrobacterium tumefaciens</i> C58	Protéobactérie Subdivision $\alpha$	5271	11	3,9	(6)
<i>Staphylococcus aureus</i>	Gram positive Bas G/C	2690	2	0,7	(7)
<i>Lactococcus lactis</i> IL403	Gram positive Bas G/C	2216	2	0,86	(8)
<i>Borrelia burgdorferi</i>	Spirochètes	1635	3	3,2	(9)
<i>Mycoplasma genitalium</i>	Gram positive Bas G/C	480	1	1,7	(10)

Tableau I-3 : Variation du nombre de facteurs  $\sigma$  putatifs dans les génomes bactériens séquencés.

\* : les ORF identifiées comme étant des transposases ont été retirées des échantillons.

(1) Bentley *et al.*, 2002 ; (2) Ikeda *et al.*, 2003 ; (3) Blattner *et al.*, 1997 ; (4) Cole *et al.*, 1998 ; (5) Kunst *et al.*, 1997 ; (6) Goodner *et al.*, 2001 ; (7) Kuroda *et al.*, 2001 ; (8) Bolotin *et al.*, 1999 ; (9) Fraser *et al.*, 1997 ; (10) Fraser *et al.*, 1995.

- CHAPITRE A-III -

L'INSTABILITE GENETIQUE AFFECTE  
L'EXPRESSION DU METABOLISME  
SECONDAIRE ET LA DIFFERENCIATION  
CELLULAIRE CHEZ LES *STREPTOMYCES*

## **A-III.1 ⌘ REARRANGEMENTS GENOMIQUES ET POLYMORPHISME INTRASPECIFIQUE CHEZ LES *STREPTOMYCES***

### **A-III.1.1 - ORGANISATION GENETIQUE ORIGINALE DU CHROMOSOME**

La linéarité du chromosome des *Streptomyces* a été découverte chez *S. lividans* 66 (Lin *et al.*, 1993). Cette particularité a ensuite été généralisée à toutes les autres espèces étudiées au niveau chromosomique. Chez *S. ambofaciens* DSM40697, l'ADN chromosomique est une molécule linéaire de 8 Mb dont les extrémités, associées à des protéines, consistent en des répétitions terminales inversées (ou TIR) de 210 kb (Leblond *et al.*, 1996). Cette structure chromosomique, appelée invertron, est une caractéristique de tous les réplicons plasmidiques et chromosomiques identifiés chez les *Streptomyces* (Kieser *et al.*, 1992; Leblond *et al.*, 1993; Lin *et al.*, 1993; Redenbach *et al.*, 1993).

Par ailleurs, le chromosome des *Streptomyces* présente une organisation génétique originale (voir paragraphe A-II.2.3.1.2.1).

En termes historiques, des premières expériences de cartographie génétique ont révélé une densité de marqueurs génétiques faible dans les régions terminales du chromosome (Kieser *et al.*, 1992). Toutefois, ces expériences utilisaient le plus souvent des mutants de fonctions du métabolisme primaire comme l'auxotrophie, présentant un phénotype facilement sélectionnable. Ces régions ont alors été appelées silencieuses. Des études fonctionnelles ont ensuite révélé la présence de nombreux gènes dans ces régions terminales (Redenbach *et al.*, 1996; Leblond and Decaris, 1999). Chez *S. ambofaciens*, les réarrangements génomiques associés à l'instabilité entraînent des changements profonds de l'expression génique (Dary *et al.*, 1993). Ainsi, la région instable renferme des gènes exprimés ou/et régulateurs de l'expression génique.

Plus récemment, le séquençage du chromosome de *S. ont* A3(2) (Bentley *et al.*, 2002) et de *S. avermitilis* (Ikeda *et al.*, 2003) ont confirmé, lorsque la fonction des gènes était identifiée, que les gènes localisés dans les « bras chromosomiques » étaient impliqués dans la différenciation cellulaire, la production des métabolites secondaires et dans la réponse à certains stress (Tableau I-4).

Gène	Espèce	Fonction et/ou condition d'expression
<i>catA/B</i>	<i>S. ont A3(2)</i>	Catalase, réponse au stress oxydatif
<i>sco2</i>	<i>S. ont A3(2)</i>	Réponse au choc thermique froid
<i>spa2</i>	<i>S. ont A3(2)</i>	Synthèse antibiotiques, phase stationnaire
<i>sapA</i>	<i>S. ont A3(2)</i>	Protéines associées aux spores, élévation du mycélium aérien
<i>abaA</i>	<i>S. ont A3(2)</i>	Différenciation cellulaire, métabolisme secondaire
<i>celA/B</i>	<i>S. ont A3(2)</i>	Cellulase extracellulaire, métabolisme secondaire
<i>crt</i>	<i>S. ont A3(2)</i>	Synthèse de caroténoïdes, phase stationnaire
<i>xln</i>	<i>S. ont A3(2)</i>	Xylanase extracellulaire, métabolisme secondaire
<i>glgBII</i>	<i>S. ont A3(2)</i>	Synthèse du glycogène en phase stationnaire
<i>gapB</i>	<i>S. ont A3(2)</i>	Glycéraldéhyde 3 phosphate déshydrogénase (gène dupliqué)
<i>argG</i>	<i>S. ont A3(2)</i>	Argininosuccinate synthétase
<i>bxlA</i>	<i>S. ont A3(2)</i>	Xylosidase extracellulaire, métabolisme secondaire
<i>IHS/S1R</i>	<i>S. ambofaciens DSM40697</i>	Régulateur putatif, phase tardive
<i>orf4.9</i>	<i>S. ambofaciens DSM40697</i>	Chitinase extracellulaire, métabolisme secondaire
<i>AUD90</i>	<i>S. ambofaciens DSM40697</i>	Polycétone synthétase (PKS) putative
<i>hasR/L</i>	<i>S. ambofaciens DSM40697</i>	Facteurs sigma alternatifs putatifs

Tableau I-4 : Exemples de gènes localisés dans la région instable chez *S. ont A(3)2* et *S. ambofaciens DSM40697* impliqués dans la différenciation cellulaire et l'expression du métabolisme secondaire (d'après (Aubert *et al.*, 1993; Aigle *et al.*, 1996; Redenbach *et al.*, 1996; Fischer *et al.*, 1998a)).

## A-III.1.2 - L'INSTABILITE GENETIQUE ET LA MODULATION DE L'EXPRESSION GENIQUE

### A-III.1.2.1 - L'INSTABILITE GENETIQUE EST CORRELEE A LA FORMATION DE REARRANGEMENTS CHROMOSOMIQUES

Les régions terminales du chromosome de *Streptomyces* sont le siège d'un phénomène d'instabilité génétique. En effet, l'instabilité génétique, définie comme l'apparition spontanée de variants à des fréquences supérieures à  $10^{-3}$ , a été corrélée à la formation de réarrangements chromosomiques dans les régions terminales du chromosome linéaire. Ces réarrangements sont de deux types : délétions et amplifications en tandem (Demuyter *et al.*, 1991). Chez *S. ambofaciens*, ces délétions peuvent atteindre la taille de 2000 kb soit le quart du génome (Leblond *et al.*, 1991). Ce phénomène a permis de définir cette région instable comme non essentielle en conditions de laboratoire.

Les délétions et amplifications terminales, sont associées à des réarrangements du génome, tels que la circularisation, des translocations chromosomiques voire même des fusions de chromosomes.

### A-III.1.2.2 - IMPACT DE L'INSTABILITE GENOMIQUE SUR L'EXPRESSION GENIQUE

#### A-III.1.2.2.1 - Modulation du dosage génique

##### A-III.1.2.2.1.1 - Amplification d'ADN

Chez *S. ambofaciens*, la surreprésentation de l'AUD90, résultant de son amplification au locus AUD90, est associée à la surexpression de *orfPS*, homologue de gènes de polycétone synthétases (PKS). La surexpression conduirait à une dérivation des précurseurs de la synthèse de la spiramycine et donc à un phénotype non producteur (Dary *et al.*, 1992), et à l'accumulation d'un nouveau métabolite secondaire de rôle inconnu (Schauner *et al.*, 1999).

De même, l'expression des gènes *spa2* et *spaR* localisés au locus amplifiable AUD6 serait modifiée chez les souches amplifiées pour ce locus. Alors que *spa2* est exprimé en phase stationnaire chez la souche sauvage, son expression n'est pas détectée chez la souche amplifiée. Par contre, *spaR* est surexprimé chez le mutant. Le gène *spa2* code un régulateur putatif homologue du régulateur (RspA) de l'expression du facteur  $\sigma$  spécifique de la phase stationnaire chez *E. coli* (RpoS). SpaR régulerait l'expression de *spa2* (Aubert *et al.*, 1993; Schneider *et al.*, 1993; Huisman and Kolter, 1994; Aubert *et al.*, 1997).

##### A-III.1.2.2.1.2 - Variation intraspécifique de la taille des répétitions terminales

Les gènes localisés dans les répétitions terminales inversées du chromosome de *S. ambofaciens* sont présents en double exemplaire. Un nouveau cluster de gènes codant une polycétone synthétase (PKS) a récemment été caractérisé au laboratoire (Pang *et al.*, sous presse). Ce cluster, nommé *alp*, pour « angucycline-like polyketide », est localisé dans les TIR du chromosome de *S. ambofaciens* ATCC2877. La construction de simples et doubles mutants de chacune des deux copies de ce cluster a permis de conclure quant à la fonctionnalité et donc l'expression de chaque cluster. Ainsi, l'expression des gènes localisés dans les répétitions terminales inversées du chromosome de *Streptomyces* pourrait être surdosée, de par la présence de ces gènes en double exemplaire.

De plus, chez *S. ambofaciens*, un polymorphisme intraspécifique de la taille des TIR a été caractérisé (Fischer *et al.*, 1998b). La taille des TIR et donc la quantité de gènes dupliqués diffèrent en fonction des réarrangements.

La délétion des extrémités chromosomiques est associée à la circularisation du chromosome (Fischer *et al.*, 1997). Dans ce cas, à l'exception de séquences répétées (par exemple, loci *rrn*, *IS* ou autres gènes dupliqués), aucune région dupliquée n'est présente. L'effet sur l'expression résulte de la perte des fonctions codées par les régions terminales. D'autre part, l'amplification d'ADN est associée à la délétion des séquences localisées entre le locus amplifiable et l'extrémité du bras chromosomique. Ainsi, le chromosome reste linéaire mais ne présente plus de TIR (Figure I-14).

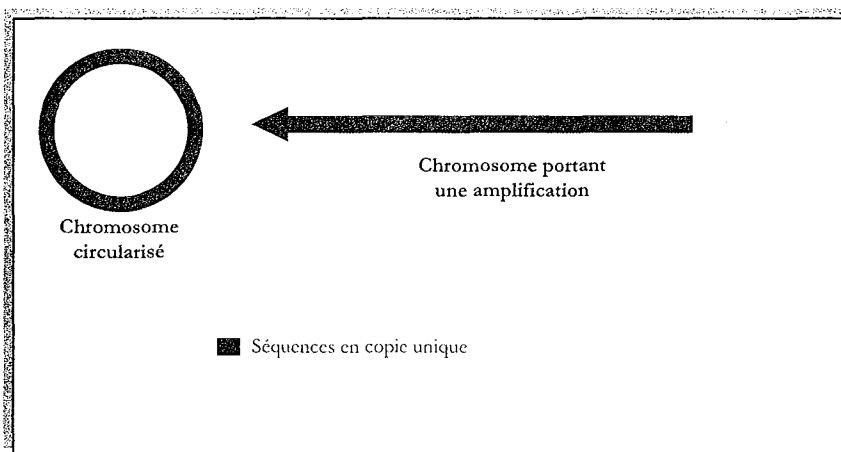


Figure I-14 : Représentation du polymorphisme de la taille des TIR observé chez *S. ambofaciens*. Cas de la perte des TIR

D'autres mutants présentent une délétion interne à un bras chromosomique entraînant la réduction de la taille des TIR, de 210 kb chez la souche sauvage à 200 kb voire 100 kb. À l'inverse, la taille des TIR augmente chez certains mutants présentant une translocation pour atteindre 850 kb (Figure I-15) (Fischer *et al.*, 1998b). Ces réarrangements résultent d'événements de recombinaison interchromosomique entre séquences dupliquées localisées à l'extérieur des répétitions terminales. Dans ce cas, la quantité de gènes dupliqués est accrue. Par exemple, dans le cas de la souche NSA855, *orfPS* codant putativement une polycétone synthétase (PKS), est dupliquée alors qu'elle se trouve en copie unique dans le chromosome sauvage.

La taille des séquences dupliquées chez *S. ambofaciens* est extrême chez les mutants présentant une fusion chromosomique. Dans ce cas, la duplication concernerait la quasi totalité de l'information génétique (environ 6.500 kb), à l'exception d'une courte région au centre de la structure (Figure I-15).



A-III.1  $\infty$  Réarrangements génomiques et polymorphisme intraspécifique chez les *Streptomyces*

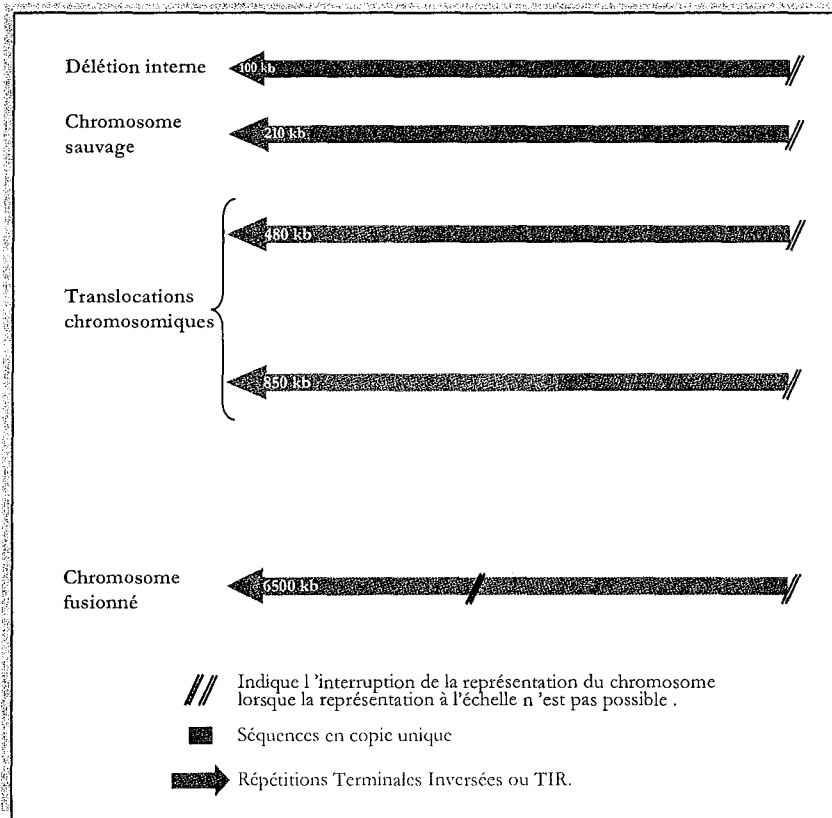


Figure I-15 : Représentation du polymorphisme de la taille des TIR observé chez *S. ambofaciens*.  
 Polymorphisme de la taille des TIR chez les mutants issus de l'instabilité génétique.

Ainsi, l'instabilité télomérique chez *S. ambofaciens* génère une grande variabilité de la quantité d'ADN dupliquée et donc de gènes dupliqués. Les chromosomes circulaires ne présentent plus de duplication. En revanche les chromosomes fusionnés présentent une duplication de plus de 6500 kb. Cette variabilité s'accompagne d'un polymorphisme intraspécifique de la taille du chromosome, d'environ 6000 kb à plus de 13000 kb.

Cette variabilité offre un modèle d'étude de l'impact des réarrangements sur l'expression génique mais, également, des mécanismes d'évolution des chromosomes par duplication.

**A-III.1.2.2.2 - Création d'un nouveau cadre d'expression des ORF *has***

Outre la duplication de grandes régions génomiques, les mutants de *S. ambofaciens* possédant notamment des translocations chromosomiques présentent, au point de cassure, une nouvelle combinaison de séquences. En effet, ces structures chromosomiques résultent d'événements de recombinaison homologue entre deux copies d'une séquence localisée à l'extérieur des répétitions terminales (Figure R-1). Ces ORF coderaient pour des facteurs  $\sigma$  alternatifs. Elles sont appelées *hasL* et *hasR* pour « Homologous to Alternative Sigma factor », L et R pour leur position

respective sur les bras chromosomiques gauche ("Left") et droit ("Right") du chromosome de la souche sauvage. Les séquences dupliquées présentent 100% d'identité en nucléotides sur 839 pb (Fischer *et al.*, 1998b).

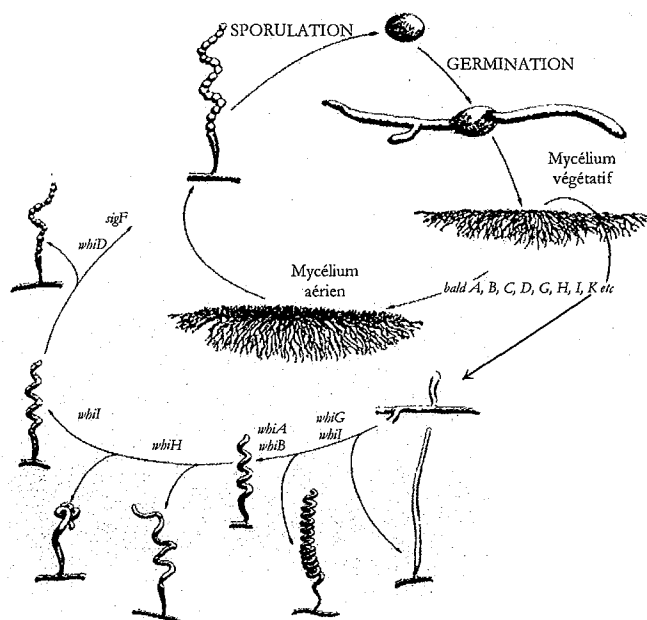
L'analyse de ce point de recombinaison a montré que des chimères des ORF *has* étaient formées au point de recombinaison (Figure R-1). Ainsi, la souche mutante NSAH135 renferme une ORF chimérique *has135*, composée de la région 5' de *hasL* et de la région 3' de *hasR*. Inversement, la souche mutante NSA855 renferme une ORF chimérique *has855* correspondant à la région 5' de *hasR* fusionnée en phase à la région 3' de *hasL*. Les copies présentes sur le bras chromosomique intact ne sont pas remaniées au cours du réarrangement. Les ORF *has* codent des polypeptides putatifs de 281 acides aminés.

## A-III.2 $\propto$ LA DIFFERENCIATION COMPLEXE DES *STREPTOMYCES* ET SON RESEAU DE REGULATION

Chez les bactéries appartenant aux genres *Streptomyces*, *Mycrococcus* et *Bacillus*, la différenciation cellulaire implique l'activation séquentielle de différents facteurs. Chaque facteur  $\sigma$  ordonne la transcription par l'ARN polymérase de gènes spécifiques. La régulation est spatio-temporelle, c'est-à-dire que l'activation des gènes dépend étroitement du stade de différenciation et de la localisation dans les différents types cellulaires.

Le développement végétatif des *Streptomyces* correspond à la germination de spores aboutissant à la formation du mycélium rampant ou végétatif. Les facteurs  $\sigma$  impliqués dans l'expression des gènes de phase végétative sont codés par les gènes *brd*, *brd* pour « homolog of *rpoD* », le gène codant le facteur  $\sigma$  principal chez *E. coli* et *B. subtilis* (Tanaka *et al.*, 1988; Lonetto *et al.*, 1994). Ces gènes sont au nombre de quatre chez *S. coelicolor* A3(2) (*brdA* à *D*). Le gène *brdB* est essentiel et code le facteur  $\sigma$  principal. Les trois autres, bien que reconnaissant des promoteurs similaires, sont non essentiels.

La différenciation cellulaire débute par la production des hyphes aériens à partir du mycélium végétatif puis à la septation du mycélium aérien en compartiments mononucléés, les spores (Figure I-16). Deux classes de mutants de différenciation ont été distinguées chez *S. coelicolor* A3(2) : les mutants Bald pour chauve (gènes *bld*) et White pour dépigmenté (gènes *whi*) (Chater, 1993). Les mutations des gènes *bld* confèrent la formation de colonies dépourvues de mycélium aérien d'où le terme chauve ("bald"). Une douzaine de gènes *bld* ont été identifiés et codent des fonctions variées : facteur de transcription pour *bldD* et *bldB*, tRNA reconnaissant un codon rare (TAA) pour *bldA*, perméase d'oligopeptide impliquée dans la transduction de stimuli extracellulaires pour *bldK* (Chater, 1998).



Selon Chater (1998)

**Figure I-16 :** Différents stades de croissance en milieu solide de *Streptomyces*. Le cycle de développement est représenté sur fond de couleur. Les différents stades de croissance du mycélium aérien associés aux mutations des gènes *whi* sont représentés sur fond gris.

Les mutants *whi* produisent des hyphes aériens incapables de former des spores matures et le pigment gris qui y est associé ; ils sont donc dépigmentés ("white"). Sept loci *whi* ont été identifiés (*whiA* à *J*). Les mutations de ces loci bloquent le développement à des stades morphologiques spécifiques de la différenciation des hyphes du mycélium aérien (Figure I-16). Les fonctions codées ne sont que partiellement élucidées : *whiB* et *whiH* codent des régulateurs de transcription (Davis and Chater, 1992; Ryding *et al.*, 1998) et *whiG* code un facteur  $\sigma$  alternatif (Chater *et al.*, 1989). En outre, le facteur  $\sigma$  alternatif extracellulaire  $\sigma^E$  (famille ECF pour "extracytoplasmique fonction sigma factors") ; (Lonetto *et al.*, 1994; Missiakas and Raina, 1998) est également impliqué dans la formation des parois cellulaires chez *S. coelicolor*.

L'expression du métabolisme secondaire est associée de façon étroite aux étapes tardives de différenciation. Un facteur  $\sigma$  alternatif appartenant à la famille des ECF, codé par le gène *sigE*, a été directement impliqué dans la régulation de la production d'actinomycine chez *S. antibioticus* (Jones *et al.*, 1997). Ainsi, la compréhension des mécanismes de régulation de l'expression génique en phases tardives de croissance revêt un intérêt industriel considérable : les *Streptomyces* produisent notamment environ 80% des antibiotiques utilisés en thérapies humaine et vétérinaire.

B - SECONDE PARTIE :

REGULATION DE L'INITIATION DE  
LA TRANSCRIPTION CHEZ LES  
BACTERIES

L'initiation de la transcription requiert, entre autres, la présence d'un facteur protéique appelé facteur sigma ( $\sigma$ ) impliqué dans la reconnaissance de la région promotrice. Ces facteurs  $\sigma$  dirigent la transcription de gènes en association avec une multitude de composants cellulaires et ce, dans un contexte génétique et environnemental spécifique. Ces facteurs constituent des éléments clés de la régulation de la différenciation biochimique et cellulaire et de la réponse aux stress notamment.

Deux mécanismes majeurs sont responsables de la régulation de l'initiation de la transcription.

Premièrement, la régulation peut s'exercer l'holoenzyme, par l'intermédiaire de molécules régulatrices se fixant à l'ADN. Aucun motif de fixation de protéine régulatrice n'ayant été détecté en amont des gènes *bas* faisant l'objet de ce travail, cette régulation ne sera pas détaillée. Très brièvement, l'ancrage du complexe enzymatique de transcription à son promoteur et la transition du complexe fermé vers un complexe ouvert (ouverture de la double hélice d'ADN) sont en effet affectés par la présence d'effecteurs protéiques, en plus de l'influence de la conformation de l'ADN au niveau de la région promotrice (pour revue, (Wosten, 1998)).

De nombreuses protéines possédant des domaines de liaison à des séquences spécifiques de l'ADN peuvent influencer l'activité promotrice. Chez *E. coli*, il existe une quarantaine de protéines impliquées dans l'activation ou la répression de la transcription à partir des promoteurs régulés par le facteur  $\sigma^{70}$  (Gralla and Collado-Vides, 1996). Ces acteurs permettent à la bactérie de réguler l'expression d'ensembles de gènes en réponse à des stimuli internes et externes définissant des régulons. Les activateurs de ces promoteurs sont habituellement liés à l'ADN entre les régions -30 à -80 du site d'initiation de la transcription mais ils peuvent également agir à des distances plus importantes. Les répresseurs se lient principalement à des positions sur l'ADN qui masquent le site de reconnaissance de l'ARN polymérase (Gralla and Collado-Vides, 1996).

Chez les *Streptomyces*, plusieurs protéines régulatrices de l'expression transcriptionnelle ont également été caractérisées. Même sans aborder cet aspect de la régulation transcriptionnelle, on ne saurait rester sans citer notamment le répresseur HspR qui, en réponse à un choc thermique, contrôle l'expression des opérons codant les chaperonnes moléculaires principales (Grandvalet *et al.*, 1997), ou encore le régulateur HrcA se fixant sur des sites nommées CIRCE (Duchene *et al.*, 1994).

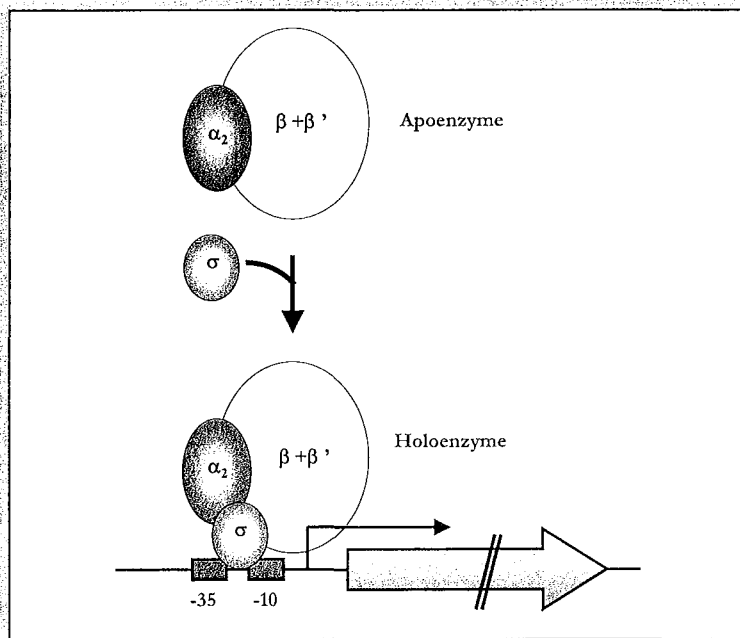
Cette partie de l'introduction se focalisera sur le second mode de régulation de l'initiation de la transcription, s'exerçant via la modulation directe de la spécificité de l'ARN polymérase par interchangeabilité du facteur sigma. A un niveau supérieur, la régulation de l'initiation de la transcription peut également s'effectuer via la régulation de l'expression des facteurs  $\sigma$  (voir Discussion).

- CHAPITRE B-I -

LE MECANISME D'INITIATION DE LA  
TRANSCRIPTION

La transcription des gènes se décompose en trois étapes : l'initiation, l'élongation et la terminaison ; l'étape d'initiation jouant un rôle fondamental dans le contrôle de l'expression des gènes.

L'initiation de la transcription bactérienne nécessite, outre l'ARN polymérase apoenzyme ( $\beta\beta'\alpha_2$ ), un facteur protéique appelé sigma ( $\sigma$ ), qui une fois associé à l'ARN polymérase forme l'holoenzyme capable alors de reconnaître une séquence particulière d'ADN localisée en amont de la région codante du gène à transcrire et appelée promoteur (classiquement composé des motifs -35 et -10 chez les promoteurs contrôlés par les facteurs de la famille  $\sigma^{70}$  d'*E. coli* ; Burgess *et al.*, 1969 ; Figure I-17).



**Figure I-17 : Mécanisme général de l'initiation de la transcription chez les bactéries.** Les motifs -35 et -10 correspondent aux régions de reconnaissance par le facteur  $\sigma$ . Le site d'initiation de la transcription ou +1 est symbolisé par la flèche en amont du cadre de lecture. La phase codante n'est pas représentée à l'échelle.

Selon leur similarité de séquences, les facteurs  $\sigma$  peuvent être regroupés en deux grandes familles : les facteurs  $\sigma$  de type  $\sigma^{70}$  et les facteurs  $\sigma$  de type  $\sigma^{54}$ . La majorité des facteurs  $\sigma$  bactériens appartiennent à la famille des facteurs  $\sigma^{70}$ , définie à partir de la séquence du facteur  $\sigma$  principal de 70 kDa d'*E. coli* (Lonetto *et al.*, 1994). Toutes les bactéries contiennent donc un ou plusieurs facteurs  $\sigma$  de type  $\sigma^{70}$  et la plupart d'entre elles codent également un facteur  $\sigma$  appartenant à la famille des facteurs  $\sigma^{54}$  (Merrick, 1993), famille définie par rapport à la séquence du facteur  $\sigma$  de 54 kDa d'*E. coli*, impliqué dans le métabolisme de l'azote.

Mise à part cette distinction structurale, les facteurs  $\sigma$  peuvent être regroupés selon les fonctions qu'ils contrôlent. En effet, certains facteurs  $\sigma$  sont dits principaux car impliqués dans la régulation de gènes dits « essentiels », alors que d'autres sont qualifiés d'alternatifs car contrôlant la

transcription de régulons spécifiques en réponse à des conditions physiologiques ou environnementales particulières.

Au sein des facteurs  $\sigma$  de type  $\sigma^{70}$ , plusieurs sous-groupes peuvent être distingués en fonction de leurs spécificités fonctionnelles et structurales (détaillées dans le chapitre B-II, Figure I-18).

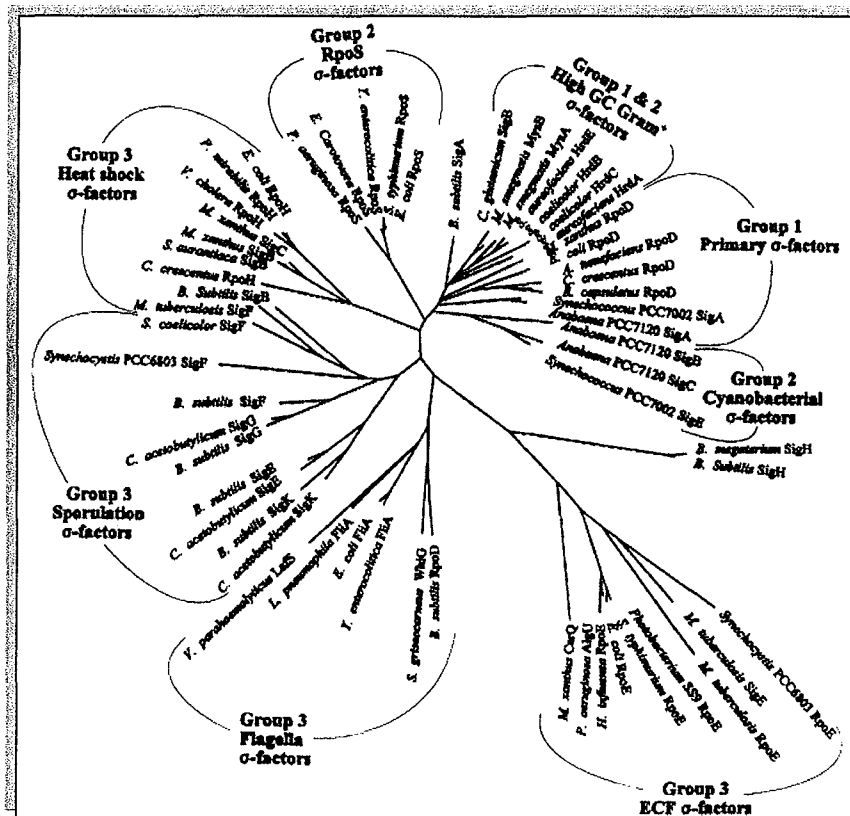


Figure I-18 : Les facteurs  $\sigma^{70}$  chez les bactéries  
Selon (Wosten, 1998). Cet arbre a été construit par la méthode FITCH basée sur les distances, à l'aide des séquences protéiques des différents facteurs  $\sigma$ .

Le séquençage systématique des génomes a révélé que le nombre de facteurs  $\sigma$  variait énormément entre les différentes bactéries, puisqu'il est compris entre 1 chez *Mycoplasma genitalium* et plus de soixante chez les *Streptomyces* (voir Tableau I-3).

La redondance des facteurs sigma est particulièrement remarquable dans la lignée des bactéries Gram positives avec 17 pour *Bacillus subtilis*, 13 pour *Mycobacterium tuberculosis*, 60 pour *S. avermitilis* et 65 pour *Streptomyces coelicolor* (voir Tableau I-3).



## B-I.1 $\omega$ LE FACTEUR SIGMA : SOUS-UNITÉ DE L'ARN POLYMERASE ADN DÉPENDANTE

Ce paragraphe traite des facteurs  $\sigma$  de type  $\sigma^{70}$ , sachant que la structure des facteurs  $\sigma$  de type  $\sigma^{54}$ , le mécanisme d'initiation de la transcription qu'ils médient ainsi que les séquences promotrices qu'ils reconnaissent leur sont spécifiques (voir paragraphe B-II.2.1.6).

### B-I.1.1 - LES ACTEURS : L'ARN POLYMERASE ADN DÉPENDANTE

#### B-I.1.1.1 - LA FORME APOENZYME DE L'ARN POLYMERASE ADN DÉPENDANTE

Chez *E. coli*, l'ARN polymérase ADN dépendante sous sa forme apoenzyme est composée de quatre sous-unités : deux sous-unités  $\alpha$ , une sous-unité  $\beta$  et une sous-unité  $\beta'$  ; chacune d'entre elles portant des fonctions spécifiques (Tableau I-5).

Cet hétérotétramère se lie à l'ADN de manière non spécifique (en l'absence du facteur  $\sigma$ ) et porte l'activité catalytique permettant la formation des liaisons 5'-3' phosphodiester entre les ribonucléosides, permettant ainsi la synthèse d'une molécule d'ARN.

Sous-unité	Gène	Fonction	Références
$\alpha$	<i>rpoA</i>	Initiation de l'assemblage de l'apoenzyme par dimérisation du domaine N-terminal	Ovchinnikov <i>et al.</i> , 1977 Ishihama <i>et al.</i> , 1981 Igarashi <i>et al.</i> , 1991
$\beta$	<i>rpoB</i>	Activité catalytique de synthèse de l'ARN et activité de correction des erreurs	Ovchinnikov <i>et al.</i> , 1981
$\beta'$	<i>rpoC</i>	Formation du site catalytique en association avec la sous-unité $\beta$	Ovchinnikov <i>et al.</i> , 1982

Tableau I-5 : Caractéristiques des sous-unités constituant la forme apoenzyme de l'ARN polymérase d'*E. coli*

### B-I.1.1.2 – LA FORME HOLOENZYME DE L'ARN POLYMERASE ADN DEPENDANTE ET LES FACTEURS SIGMA PRINCIPAUX

L'apoenzyme interagit avec le facteur  $\sigma$ , reconnaissant les régions promotrices, pour former l'holoenzyme assurant ainsi l'initiation de la transcription. Différents facteurs  $\sigma$  peuvent s'associer à l'apoenzyme pour former différentes holoenzymes ( $E\sigma$ ) présentant chacune une spécificité de reconnaissance pour des régions promotrices distinctes.

#### B-I.1.1.2.1 – Les facteurs sigma principaux chez les bactéries

Les facteurs  $\sigma$  principaux reconnaissent une large majorité de promoteurs transcriptionnels de gènes impliqués dans les fonctions essentielles et assurent ainsi la transcription « de base » chez les bactéries.

Chez *E. coli*, le facteur principal est appelé  $\sigma^{70}$  et est codé par le gène *rpoD* (Burton *et al.*, 1981).

Chez *B. subtilis*, ce facteur sigma principal est appelé  $\sigma^A$  (Shorenstein and Losick, 1973) mais est également impliqué dans l'initiation de la sporulation (voir ci après).

Chez *Streptomyces coelicolor*, quatre homologues du gène *rpoD* ont été identifiés : *brdA*, *brdB*, *brdC* et *brdD* (Buttner *et al.*, 1990). Le facteur sigma codé par le gène *brdB* chez *S. coelicolor* est le facteur  $\sigma$  principal, les trois autres pouvant être considérés comme des paralogues non essentiels (Wosten, 1998). Le gène *brdD* est principalement exprimé en phase exponentielle de croissance et le facteur  $\sigma^{\text{HrdD}}$  reconnaît des promoteurs en amont de *redD* et *actII-orf4*, deux gènes régulateurs impliqués dans la synthèse d'antibiotiques (Fujii *et al.*, 1996; Kang *et al.*, 1997).

En revanche, bien que les gènes *brdA* et *brdC* sont exprimés et peuvent être interrompus, leur fonction reste inconnue (Buttner and Lewis, 1992).

#### B-I.1.1.2.2 – La structure des facteurs sigma de type $\sigma^{70}$

La comparaison des séquences des facteurs  $\sigma$  principaux de différentes espèces bactériennes, ainsi que des facteurs  $\sigma$  alternatifs qui seront décrits ultérieurement, a révélé l'existence de quatre domaines protéiques fortement conservés (Lonetto *et al.*, 1992)(Figure I-19).

- La région 1 se décompose en deux sous-régions : 1.1 et 1.2.

Elle présente une faible conservation de séquence et n'est présente que dans les homologues du facteur  $\sigma^{70}$  d' *E. coli* chez les bactéries.

En se repliant sur les régions impliquées dans la reconnaissance des séquences promotrices, les résidus contenus dans une première partie de la région 1.1 empêchent la fixation du facteur  $\sigma^{70}$  seul à l'ADN. Lors de l'interaction avec l'ARN polymérase, le facteur  $\sigma^{70}$  subit alors une modification structurale permettant l'accessibilité des régions interagissant avec l'ADN (Dombroski *et al.*, 1992). De plus une boucle de résidus acides, présente dans la région non conservée entre les domaines 1.2 et 2.1 pourrait aussi participer à l'inhibition de la fixation du facteur  $\sigma^{70}$  seul à l'ADN (Malhotra *et al.*, 1996). Les résidus contenus dans la seconde partie de la région 1.1 ainsi que dans la région 1.2 semblent impliqués dans la formation des complexes d'initiation (Wilson and Dombroski, 1997).

- **La région 2** peut être subdivisée en cinq sous-domaines : 5.1, 5.2, 5.3, 5.4 et 5.5.

Elle est fortement conservée dans les facteurs  $\sigma$  de nombreuses bactéries (Gribskov and Burgess, 1986). Les sous-domaines 2.1 et 2.2 seraient impliqués dans l'interaction du facteur  $\sigma^{70}$  avec l'ARN polymérase (Lesley and Burgess, 1989). De nombreuses études génétiques ont par ailleurs montré l'implication de la région 2.4 des facteurs  $\sigma$  dans la reconnaissance et la fixation des séquences -10 des promoteurs (Dombroski *et al.*, 1992). De plus, le facteur  $\sigma^{70}$  reconnaîtrait la séquence TG des promoteurs dits « étendus » (voir paragraphe B-I.2) par l'intermédiaire de résidus fortement conservés de la région 2.5 (Barne *et al.*, 1997). Enfin le rôle de la région 2.3 n'a pas encore été clarifié.

- **La région 3** se divise en deux sous-régions : 3.1 et 3.2.

Elle semble être impliquée, tout comme les régions 2.1 et 2.2, dans l'interaction des facteurs  $\sigma$  avec l'ARN polymérase. De plus, cette région pourrait intervenir dans le processus d'initiation de la transcription, puisque son extrémité carboxy-terminale participe au site d'initiation de la transcription de l'holoenzyme (Severinov *et al.*, 1994).

- **La région 4** se décompose en deux sous-domaines : 4.1 et 4.2.

Les sous-domaines 4.1 et 4.2 semblent être impliqués dans l'interaction du facteur  $\sigma^{70}$  avec les activateurs transcriptionnels. De plus, la sous-région 4.2 permet la reconnaissance spécifique de la séquence -35 des promoteurs par l'intermédiaire d'une structure de type hélice-coude-hélice (HTH) (Keener and Nomura, 1993).

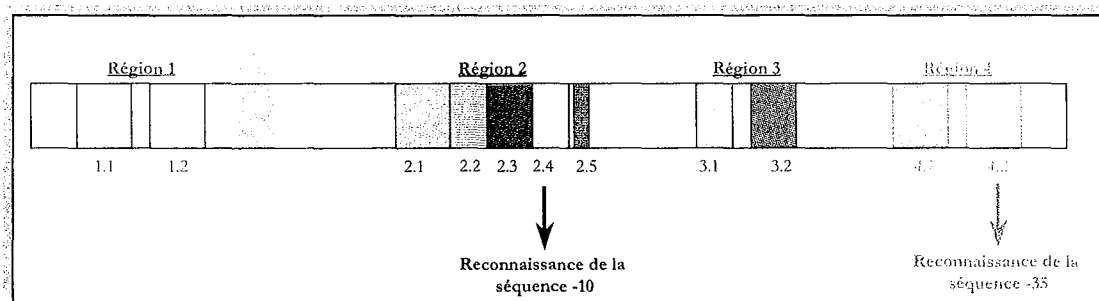


Figure I-19 : Structure du facteur  $\sigma^{70}$  d'*E. coli*

Les différentes régions conservées et leurs fonctions dans l'initiation de la transcription sont indiquées.

## **B-I.1.2 - LE MECANISME DE L'INITIATION DE LA TRANSCRIPTION**

Les études sur le mécanisme de l'initiation de la transcription ont été principalement réalisées à partir de promoteurs reconnus par l'holoenzyme  $E\sigma^{70}$  (pour revue, (Lonetto *et al.*, 1992) ; Figure I-20).

L'initiation de la transcription est un processus nécessitant la formation de complexes successifs ARN polymérase / promoteur, aboutissant à la progression de l'ARN polymérase le long de l'ADN avec élongation de la chaîne d'ARN.

Cette synthèse d'une molécule d'ARN, étant effectuée par copie du brin matrice de la double hélice d'ADN appelé brin codant, nécessite donc une séparation locale des deux brins d'ADN.

### **□ Le complexe fermé RPc**

Dans une première étape, l'holoenzyme, fixée dans un premier temps à l'ADN de façon non spécifique, se déplacerait sur l'ADN puis reconnaîtrait une région promotrice, par interaction du facteur  $\sigma^{70}$  avec la région -35 puis avec la région -10. Le complexe qui est alors formé, RPc pour « ARN polymérase / promoter closed complex », est qualifié de fermé puisque à cette étape les deux brins d'ADN ne sont pas encore séparés.

### **□ Le complexe intermédiaire RPint**

La seconde étape de l'initiation transcriptionnelle conduit à la formation d'un complexe intermédiaire RPint. Dans ce complexe, la région d'ADN située entre les boîtes -10 et -35 serait déroulée afin que ces deux boîtes se retrouvent alignées sur une même face de la double hélice d'ADN (pour revue, (deHaseth and Helmann, 1995)). De plus, la formation de ce complexe intermédiaire implique des réarrangements structuraux de l'ARN polymérase (Sen and Dasgupta, 1996).

### **□ Le complexe ouvert RPo**

Les fortes contraintes structurales imposées à l'ADN lors de la formation du complexe intermédiaire RPint, induisent alors l'ouverture de la double hélice d'ADN avec formation du complexe ouvert RPo (pour « ARN polymérase / promoter open complex »).

Dans le complexe RPo, la région d'ADN comprise entre les positions -9 et +3 est sous forme simple brin.

### **□ Le complexe d'initiation RPinit**

Le site +1 d'initiation de la transcription est alors accessible au ribonucléotide initiateur et l'ARN polymérase initie la polymérisation des premiers ribonucléotides au sein du complexe d'initiation nommé RPinit.

□ Le complexe d'élongation RPElong

Après polymérisation de 8 à 10 ribonucléotides, le facteur  $\sigma$  se dissocie de l'ARN polymérase qui, sous sa forme apoenzyme, va se déplacer le long de l'ADN et passer en phase d'élongation (RPElong). Une molécule d'ARN sera alors synthétisée jusqu'à ce que l'ARN polymérase rencontre un signal de terminaison de la transcription.

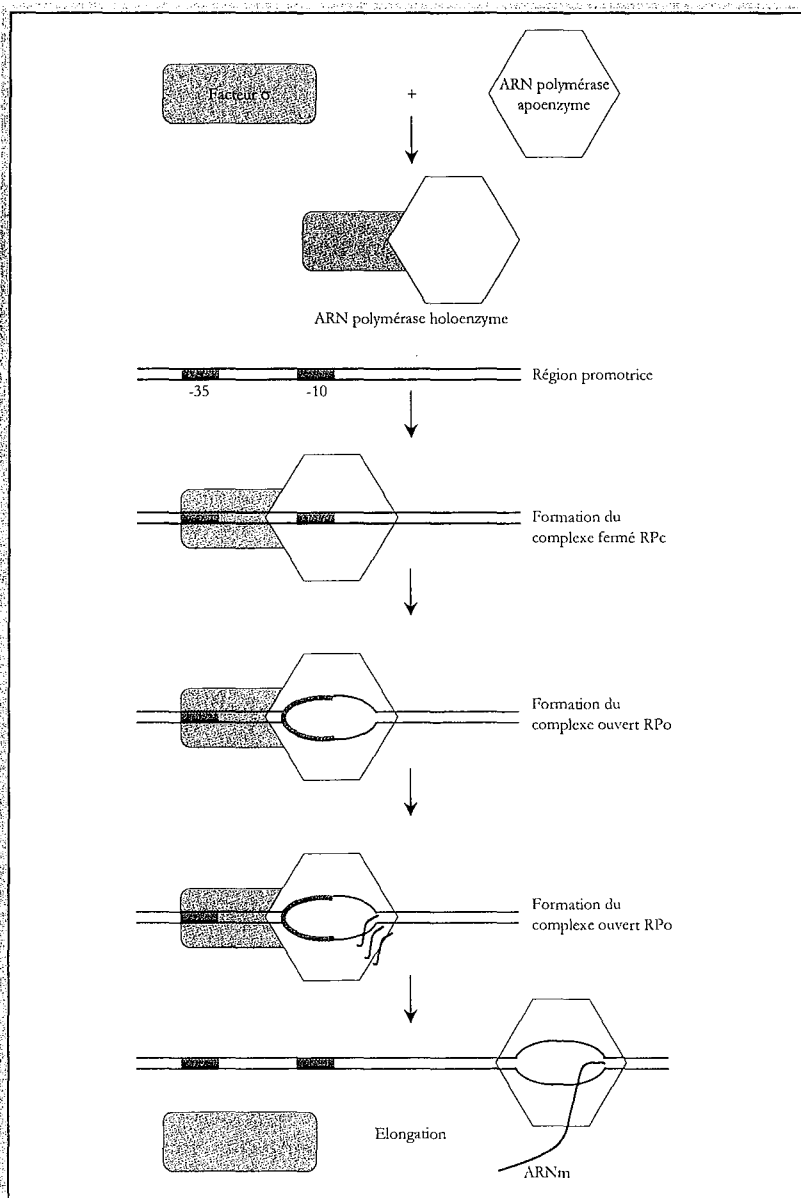
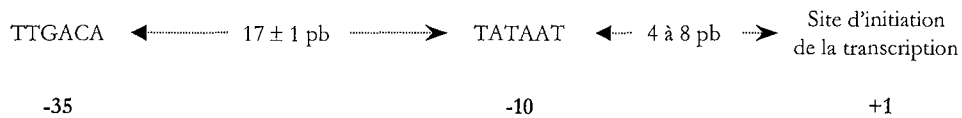


Figure I-20 : Les différentes étapes de l'initiation de la transcription chez *E. coli* d'après Polyakov *et al.* (1995).

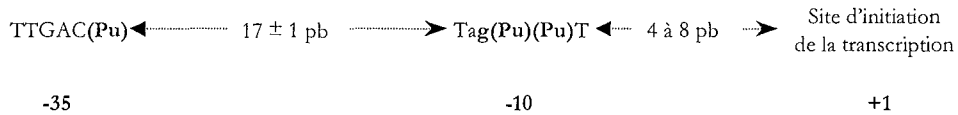
## B-I.2 ☞ LES SEQUENCES PROMOTRICES RECONNUS PAR LES FACTEURS SIGMA DE TYPE $\sigma^{70}$

Le facteur  $\sigma^{70}$  permet la reconnaissance d'une séquence d'ADN appelée promoteur de type facteur  $\sigma^{70}$ . L'alignement des séquences situées en amont du site d'initiation de la transcription des gènes de ménage d'*E. coli* a permis de définir deux motifs conservés (Harley and Reynolds, 1987) et donc d'obtenir le consensus suivant :



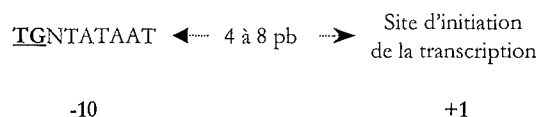
La séquence non conservée entre les motifs  $-10$  et  $-35$  varie de 16 à 18 pb, la distance de 17 pb étant préférentiellement observée. Cette séquence consensus est fortement conservée dans le monde bactérien (pour revue, (Helmann and Chamberlin, 1988)). Dans le cas de *Bacillus subtilis*, la séquence consensus reconnue par le facteur sigma principal  $\sigma^A$  est totalement similaire à celle d'*E. coli*.

En revanche, chez les *Streptomyces*, la séquence reconnue par le facteur sigma principal  $\sigma^{\text{HrdB}}$  diverge légèrement, vraisemblablement à cause du fort pourcentage en bases G et C de leur génome (Strohl, 1992) :



Outre les promoteurs contenant une séquence  $-10$  et  $-35$ , certains promoteurs, appelés « promoteurs étendus » présentent la particularité de posséder un dinucléotide TG en amont de la boîte  $-10$ . La séquence  $-35$  de ces promoteurs n'est pas essentielle à la reconnaissance par l'holoenzyme mais peut cependant influencer leur efficacité (Kumar *et al.*, 1993).

La séquence consensus de ces promoteurs est donc la suivante :



Ce motif  $-10$  « étendu » est également fortement conservé dans le monde bactérien. Des promoteurs « étendus » ont été identifiés chez *Clostridium pasteurianum* (Graves and Rabinowitz, 1986), chez *Streptococcus pneumoniae* (Sabelnikov *et al.*, 1995), ou encore chez différentes espèces de *Mycobacterium* (Bashyam and Tyagi, 1998). Chez *B. subtilis*, ce motif  $-10$  « étendu » est fréquemment rencontré : en effet 45% des promoteurs identifiés chez cette bactérie possèdent ce dinucléotide TG (Helmann, 1995).

Chez toutes ces bactéries aucune corrélation n'a pu être faite entre présence de promoteurs « étendus » et fonction des gènes.

- CHAPITRE B-II -

**REGULATION TRANSCRIPTIONNELLE  
ET FACTEURS SIGMA ALTERNATIFS**

Les évidences de la reprogrammation de l'ARN Polymérase (ARNP) par des  $\sigma$  distincts sont issues de l'étude de cellules infectées par des bactériophages (Fox and Pero, 1974; Kassavetis and Geiduschek, 1984). Les bactériophages SPO1 et SP82 de *Bacillus subtilis* ainsi que le phage T4 d'*E. coli* codent des protéines régulatrices capables de diriger la transcription de régions spécifiques de leur génome par le biais de protéines hôtes. Les premiers facteurs  $\sigma$  "alternatifs" ont de surcroît été purifiés à partir de cellules de *B. subtilis* infectées par SPO1. Ceux-ci peuvent s'associer avec l'apoenzyme hôte et, en absence du facteur  $\sigma$  principal  $\sigma^A$  de *B. subtilis*, entraîner l'initiation de la transcription à partir de promoteurs phagiques. La découverte de multiples facteurs  $\sigma$  au sein de bactéries infectées par des bactériophages établit un précédent dans la notion de contrôle de l'expression génique. Depuis, l'étude des phénomènes de différenciation cellulaire chez les bactéries, et notamment la sporulation chez *B. subtilis*, ou encore la réponse aux stress chez *E. coli*, ont généralisé ce concept aux processus de développement.

## B-II.1 $\sigma$ PRÉSENTATION DES FACTEURS SIGMA ALTERNATIFS

Les facteurs sigma alternatifs contrôlent la transcription de régulons spécifiques en réponse à des conditions physiologiques ou environnementales particulières.

Différentes sous-familles peuvent être distinguées au sein des facteurs  $\sigma$  alternatifs, notamment selon leur fonction (Figure I-18) :

- ❑ Les facteurs  $\sigma$  impliqués dans la biosynthèse des flagelles,
- ❑ Les facteurs  $\sigma$  à fonction extracytoplasmique ou ECF,
- ❑ Les facteurs  $\sigma$  de réponse au choc thermique,
- ❑ Les facteurs  $\sigma$  impliqués dans la sporulation.

Les facteurs  $\sigma$  à fonction extracytoplasmique ou ECF vont être brièvement présentés dans cette partie, les autres types de facteurs  $\sigma$  alternatifs étant détaillés par la suite chez les trois organismes *E. coli*, *B. subtilis* et les *Streptomyces* (voir paragraphe B-II.2).

Les facteurs  $\sigma$  qui appartiennent à la famille des ECF (famille ECF pour "extracytoplasmic function sigma factors" ; (Lonetto *et al.*, 1994; Missiakas and Raina, 1998)) régulent une large variété de procédés en relation avec l'environnement ont, tels que la production d'alginate (Deretic *et al.*, 1994), le transport de citrate ferrique (Angerer *et al.*, 1995) ou la synthèse des caroténoïdes (Gorham *et al.*, 1996).



Les différentes caractéristiques qui distinguent les facteurs  $\sigma$  de la famille des ECF des autres membres de la famille des facteurs  $\sigma$  de type  $\sigma^{70}$  sont les suivantes :

- l'absence des domaines protéiques conservés 1 et 3 et donc leur petite taille (classiquement 20 à 30 kDa),
- la reconnaissance de séquences promotrices proches,
- leur fréquente régulation post-traductionnelle par anti-facteur  $\sigma$  spécifique, généralement codé par un gène situé en aval du gène codant le facteur  $\sigma$ .

Grâce au séquençage systématique des génomes, de nombreux facteurs  $\sigma$  de la famille des ECF ont pu être identifiés mais leur fonction reste souvent inconnue. Par exemple, 45 et 47 facteurs  $\sigma$  de type ECF ont été identifiés dans les génomes des deux espèces de *Streptomyces* séquencés (voir paragraphe A-II.2.3.2.3, (Bentley *et al.*, 2002; Ikeda *et al.*, 2003)). L'analyse du génome de *B. subtilis* a révélé l'existence de sept facteurs  $\sigma$  de la famille des ECF (Kunst *et al.*, 1997). Aucun d'entre eux n'avait précédemment été découvert par des approches de génétique classique ; ceci pouvant être expliqué par le fait que les facteurs  $\sigma$  de la famille des ECF régulent souvent des gènes transcrits par plus d'un promoteur, impliquant donc que la mutation nulle de ces facteurs  $\sigma$  n'entraînent pas l'extinction complète de la transcription du gène cible (Paget *et al.*, 1998; Paget *et al.*, 1999).

## B-II.2 $\propto$ LES FACTEURS SIGMA ALTERNATIFS ET LA COMPLEXITE DE REPONSE A UN STRESS ENVIRONNEMENTAL D'UN ORGANISME

### B-II.2.1 – LES FACTEURS SIGMA ALTERNATIFS CHEZ *ESCHERICHIA COLI*

Chez *E. coli* six facteurs  $\sigma$  distincts du facteur  $\sigma^{70}$  ont été mis en évidence. Pour chacun d'entre eux une séquence consensus a pu être dégagée à partir de l'alignement des séquences nucléotidiques de chaque type de promoteur (Tableau I-6).

Ces facteurs  $\sigma$  peuvent être divisés en deux groupes (pour revue (Lonetto *et al.*, 1992)). En effet la plupart des facteurs sigma alternatifs présentent de fortes homologues avec le facteur  $\sigma^{70}$  et le mécanisme d'initiation de la transcription pour ces holoenzymes est proche de celui décrit pour l'holoenzyme  $E\sigma^{70}$ . Une exception concerne le facteur  $\sigma^{54}$ , impliqué dans la transcription des gènes du métabolisme de l'azote, qui permet une initiation de la transcription par un mécanisme plus proche de celui observé chez les eucaryotes.

$\sigma$	Gène	PM (Da)	Promoteur		Fonction	Réf	
			-35	-10			
$\sigma^{70}$	<i>rpoD</i>	70236	TTGACA	N16-18	TATAAT	Majorité des gènes d' <i>E. coli</i>	(1)
$\sigma^{38}$ (S)	<i>rpoS</i>	38000	CTGCAA	N14-16 *	CGCCAAGT	Réponse à de nombreux stress	(2) (3)
$\sigma^{32}$	<i>rpoH</i>	32000	CTTGAA	N13-15	CCCCATNT A	Réponse à la dénaturation des protéines périplasmiques	(4)
$\sigma^{24}$ (E)	<i>rpoE</i>	24000	GAACCTT	N16	TCTGA	Réponse à la dénaturation des protéines périplasmiques	(5)
$\sigma^{28}$ (F)	<i>rpoF</i>	28000	TAAACTT	N11	GCCGATAA	Gènes de synthèse des flagelles	(6)
$\sigma^{54}$ (N)	<i>rpoN</i>	54000	GTGGCAC -26	N5	TTGC -12	Gènes du métabolisme de l'azote	(7)
$\sigma^{19}$	<i>fecI</i>	19000	GGAAAAT -35	N17	TCCTTT -10	Opéron de transport du citrate ferrique	(8)

Tableau I-6 : Les différents facteurs  $\sigma$  mis en évidence chez *E. coli*.

Pour chaque facteur  $\sigma$ , le nom de son gène, son poids moléculaire, la séquence reconnue par l'holoenzyme correspondante ainsi que le type de gènes transcrits par cette holoenzyme sont indiqués.

\* : Le facteur  $\sigma^{38}$  reconnaît également des séquences de type  $\sigma^{70}$ .

(1) Harley et Reynolds, 1987 ; (2) Lange et Hengge-Aronis, 1991 ; (3) Kolb *et al.*, 1995 ; Tanaka *et al.*, 1995 ; (4) Cowing *et al.*, 1985 ; (5) Erickson et Gross, 1989 ; (6) Helmann et Chamberlin, 1987 ; (7) Hunt et Magasanik, 1985 ; (8) Angerer *et al.*, 1995 ; ENZ *et al.*, 1995.

### B-II.2.1.1 – LE FACTEUR $\sigma^{32}$ ET LA REPONSE AU CHOC THERMIQUE

Le facteur  $\sigma^{32}$  est impliqué dans la réponse à différents stress conduisant à la production de protéines dénaturées dans le cytoplasme, tel que le choc thermique (Cowing *et al.*, 1985). Il permet en particulier la transcription des gènes codant les protéines chaperons (pour revue, (Mager and De Kruijff, 1995) ; Tableau I-6).

La transcription du gène *rpoH* est activée en conditions de stress. En effet, ce gène est alors transcrit à partir du promoteur P3 par l'holoenzyme alternative E $\sigma^{24}$  décrite ci-après (pour revue, (Mager and De Kruijff, 1995)).

La quantité de facteur  $\sigma^{32}$  présent dans la cellule est également régulée aux niveaux des étapes d'initiation de la traduction et de dégradation protéique. En effet, la fixation des protéines chaperons DnaJ et DnaK-GrpE sur le facteur  $\sigma^{32}$  inhibe sa liaison à l'ARN polymérase et favorise sa dégradation par des protéases. Lors d'un choc thermique, les chaperonnes ne se fixeront plus sur le facteur  $\sigma^{32}$  mais sur les nombreuses protéines cytoplasmiques dénaturées, entraînant alors la stabilité de ce facteur  $\sigma$ . De plus, cette régulation permet un rétrocontrôle, puisque le facteur  $\sigma^{32}$  est lui même responsable de la transcription des protéines chaperons inhibant son action (pour revue, (Mager and De Kruijff, 1995)).

### B-II.2.1.2 – LE FACTEUR $\sigma^{24}$ OU $\sigma^E$ ET LE CHOC THERMIQUE PROLONGE

Le facteur  $\sigma^{24}$  ou  $\sigma^E$  permet d'induire une réponse cellulaire à des conditions de stress menaçant l'intégrité cellulaire au niveau périplasmique, tels les chocs thermiques prolongés (Connolly *et al.*, 1997).

L'holoenzyme  $E\sigma^{24}$  reconnaît des séquences -10 et -35 fortement divergentes du consensus des promoteurs reconnus par le facteur  $\sigma^{70}$  (Tableau I-6) et est notamment responsable de la transcription du gène *rpoH* codant le facteur  $\sigma^{32}$ .

L'activité du facteur  $\sigma^{24}$  ou  $\sigma^E$  est régulée par la fixation d'un anti-facteur  $\sigma$ , nommé RseA et codé par un gène localisé dans le même opéron que le gène *rpoE* (Missiakas *et al.*, 1997).

### B-II.2.1.3 – LE FACTEUR $\sigma^{28}$ OU $\sigma^F$ ET LA MOBILITE

Le facteur  $\sigma^{28}$  ou  $\sigma^F$  permet la transcription des gènes impliqués dans la mobilité des bactéries, dont les gènes tardifs nécessaires à la formation des flagelles (Tableau I-6).

Son activité est également régulée par un anti-facteur  $\sigma$ , nommé FlgM (Kutsukake and Iino, 1994; Kutsukake *et al.*, 1994).

### B-II.2.1.4 – LE FACTEUR $\sigma^{38}$ OU $\sigma^S$ ET L'ENTREE EN PHASE STATIONNAIRE

Le facteur  $\sigma^{38}$  ou  $\sigma^S$  est produit en réponse à de nombreuses conditions de stress, et notamment lors de la phase stationnaire de croissance (Hengge-Aronis, 1993), des chocs thermiques (Muffler *et al.*, 1997) ou osmotiques (Hengge-Aronis, 1996).

L'holoenzyme  $E\sigma^{38}$  est capable de reconnaître deux types de séquences promotrices : des promoteurs spécifiques à  $\sigma^{38}$  (décrits dans le Tableau I-6) et les promoteurs reconnus également par le facteur  $\sigma^{70}$  (Kolb *et al.*, 1995; Tanaka *et al.*, 1995).

Comme pour le facteur  $\sigma^{32}$ , l'initiation de la transcription du gène *rpoS*, l'efficacité de la traduction ou la stabilité de la protéine  $\sigma^{38}$  peuvent être modifiées en fonction des stress cellulaires par des mécanismes complexes (Lange and Hengge-Aronis, 1994).

De plus, un niveau de régulation supplémentaire permet de compenser la plus faible concentration cellulaire et la plus faible affinité pour l'ARN polymérase du facteur  $\sigma^{38}$  par rapport au facteur  $\sigma^{70}$ , suite à l'inhibition spécifique du facteur  $\sigma^{70}$  par une protéine nommée Rsd, jouant un rôle d'anti-facteur  $\sigma$  (Jishage and Ishihama, 1998).

### B-II.2.1.5 – LE FACTEUR $\sigma^{19}$ OU PROTEINE FecI ET LE TRANSPORT DU CITRATE FERRIQUE

Le facteur  $\sigma^{19}$  ou protéine FecI permet la transcription de l'opéron *fec*, impliqué dans le transport du citrate ferrique (Tableau I-6 ; (Angerer *et al.*, 1995; Enz *et al.*, 1995)), en présence de la protéine FecR, codée par un gène de cet opéron (Ochs *et al.*, 1995).

L'activité cytoplasmique du facteur  $\sigma^{19}$  est régulée par la concentration extracellulaire de citrate ferrique via un système de transduction du signal impliquant la protéine de transport extra-membranaire FecA, le complexe Ton, et la protéine membranaire FecR (Harle *et al.*, 1995).

### B-II.2.1.6 – LE FACTEUR $\sigma^{54}$ OU $\sigma^N$ ET LE METABOLISME DE L'AZOTE

Le facteur  $\sigma^{54}$  ou  $\sigma^N$  permet la transcription de gènes impliqués dans le métabolisme de l'azote (Tableau I-6 ; (Sasse-Dwight and Gralla, 1988)).

La structure de ce facteur  $\sigma$  diffère fortement de celle des facteurs  $\sigma$  décrits précédemment, puisque semble être composée de cinq domaines structuraux interagissant entre eux (Cannon *et al.*, 1995). Les séquences promotrices reconnues par le facteur  $\sigma^{54}$  sont également très différentes, puisque sont centrées aux positions -12 et -24 par rapport au site +1 d'initiation de la transcription (Hunt and Magasanik, 1985).

Enfin, le mécanisme d'initiation de la transcription diffère du mécanisme classique décrit précédemment et ressemble plus à celui connu chez les eucaryotes. En effet, l'holoenzyme  $E\sigma^{54}$  se fixe sur les promoteurs pour former un complexe fermé qui restera dans un état inactif jusqu'à ce que l'azote devienne limitant. Alors une cascade de réactions de phosphorylation permettra la fixation d'une protéine, nommée NtrC, qui interagira avec le complexe fermé afin d'activer la transcription (Sasse-Dwight and Gralla, 1988); (Keener and Kustu, 1988).

Ces facteurs  $\sigma$  alternatifs semblent être conservés dans de nombreuses bactéries à Gram négatif, notamment chez *Vibrio cholerae* (Klose and Mekalanos, 1998).

## **B-II.2.2 – LES FACTEURS SIGMA ALTERNATIFS CHEZ *BACILLUS SUBTILIS***

Chez *B. subtilis*, 17 facteurs  $\sigma$  différents ont pu être identifiés et au moins six d'entre eux contrôlent la transcription des gènes spécifiques de la sporulation.

### B-II.2.2.1 – LA SPORULATION CHEZ *BACILLUS SUBTILIS*

La Figure I-21 schématise l'activation séquentielle de ces facteurs  $\sigma$  (aussi bien en termes temporels qu'en termes de localisation) et les principaux changements morphologiques entraînés par chaque nouveau profil d'expression génique.

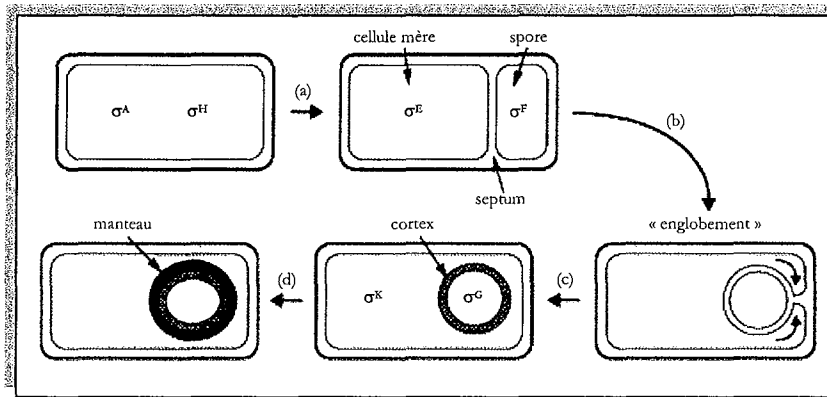


Figure I-21 : Cascade de facteurs  $\sigma$  durant la sporulation de *B. subtilis* (d'après (Kroos and Yu, 2000)).

- (a) L'expression des gènes sous le contrôle des facteurs  $\sigma^A$  et  $\sigma^H$  entraîne la formation d'un septum, en position asymétrique.
- (b) Le facteur  $\sigma^F$  devient actif dans la spore et le facteur  $\sigma^E$  dans la cellule mère. Le septum migre puis, par un procédé proche de la phagocytose, provoque « l'englobement » de la spore dans une double membrane.
- (c) Le facteur  $\sigma^G$  devient actif dans la spore, générant ainsi un signal qui active le facteur  $\sigma^K$  dans la cellule mère. Le cortex est alors synthétisé entre les deux membranes entourant la spore.
- (d) Les produits des gènes transcrits par le facteur  $\sigma^K$ , ainsi que les protéines dépendantes de  $\sigma^E$  synthétisées plus tôt dans la cellule mère, s'assemblent à la surface de la spore et produisent ainsi le manteau.

Différents mécanismes de régulation contrôlent l'activation séquentielle de chacun de ces facteurs  $\sigma$  impliqués dans la sporulation (pour revue, voir (Kroos and Yu, 2000)).

Grâce à un système de phospho-relai, un régulateur de l'initiation de la sporulation se fixe sur les régions promotrices et activent la transcription médiée par les facteurs  $\sigma^A$  et  $\sigma^H$ . Ce régulateur, appelé SpoOA-P, exercerait cette activation en se liant à la fois au facteur  $\sigma$  et au core enzyme de l'ARN polymérase (Grossman, 1995). Le niveau et l'activité du facteur  $\sigma^H$  sont contrôlés par des mécanismes transcriptionnels et post-traductionnels. La transcription du gène *spoOH* est activée suite à l'inhibition, par SpoOA-P, de la production de AbrB, un répresseur de *spoOH*. De plus, l'activité du facteur  $\sigma^H$  est influencée par la sous-unité ATPasique ClpX de la protéase ClpXP (Liu *et al.*, 1999).

Le facteur  $\sigma^F$  est régulé par un système anti-anti- $\sigma$ /anti- $\sigma$  codé par les gènes SpoIIAA/SpoIIAB (pour revue, (Helmann, 1999)), de façon similaire au facteur  $\sigma^B$  impliqué dans la réponse générale aux stress (voir paragraphe B-II.2.2.2 ci-après). La localisation de la phosphatase SpoIIE, impliquée dans ce système, à proximité du septum pourrait révéler un probable lien entre la formation du septum et l'activation du facteur  $\sigma^F$  dans la spore.

Le facteur  $\sigma^E$  est synthétisé sous forme d'un précurseur inactif,  $\text{pro-}\sigma^E$ , présentant une extension N-terminale de 27 acides aminés. La maturation de  $\text{pro-}\sigma^E$  est assurée par SpoIIR, le produit d'un gène contrôlé par le facteur  $\sigma^F$  (Stragier and Losick, 1996).

Le facteur  $\sigma^F$  est responsable de la transcription du gène codant le facteur  $\sigma^G$  dans la spore. De plus, il semblerait que SpoIIAB, l'anti- $\sigma$  du facteur  $\sigma^F$ , inhibe également l'activité du facteur  $\sigma^G$ . L'étape de transcription ainsi que la dissociation de l'anti- $\sigma$  sont dépendantes de l'activité du facteur  $\sigma^E$  dans la cellule mère (pour revue, (Kross *et al.*, 1999)).

Comme pour le facteur  $\sigma^E$ , le facteur  $\sigma^K$  est synthétisé sous forme d'un  $\text{pro-}\sigma^K$ . Le réarrangement créant le gène du  $\text{pro-}\sigma^K$  se déroulant exclusivement dans la cellule mère (voir paragraphe A-I.1.1), et ce gène étant sous le contrôle du facteur  $\sigma^E$ , la synthèse du  $\text{pro-}\sigma^K$  est confinée dans la cellule mère. Au moins deux mécanismes distincts sont responsables de la transition de la transcription médiée par le facteur  $\sigma^E$  vers celle médiée par le facteur  $\sigma^K$  dans la cellule mère. En effet, ces deux facteurs  $\sigma$  entreraient en compétition pour la fixation au core enzyme. De plus, il semblerait que le facteur  $\sigma^K$  soit responsable d'une inhibition de la transcription du gène codant le  $\text{pro-}\sigma^E$  ainsi que d'autres gènes d'expression précoce (Zhang *et al.*, 1999).

L'ensemble de ces données démontre la complexité et l'interconnexion des systèmes de régulation des facteurs  $\sigma$ , aboutissant à la sporulation chez *B. subtilis* (Figure I-22)

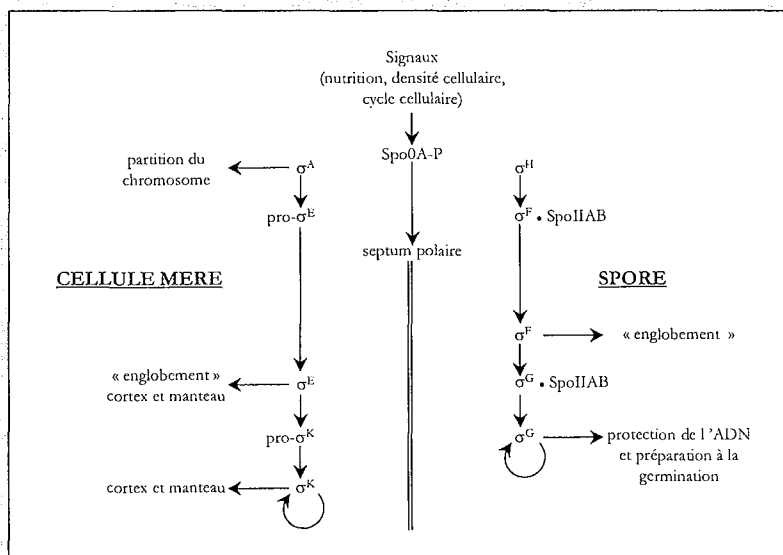


Figure I-22 : Réseau de régulation contrôlant la synthèse des facteurs  $\sigma$  et leur activation.

Les flèches continues indiquent des relations de dépendance entre les facteurs  $\sigma$  ou la relation entre le facteur  $\sigma$  et le changement morphologique engendré.

Les deux lignes verticales symbolisent les membranes qui séparent la cellule mère de la spore après la formation du septum polaire.

Les flèches en pointillés représentent les interactions par transduction du signal entre les deux types de cellules qui gouvernent l'activation séquentielle des facteurs  $\sigma$  (petits pointillés) ou leur synthèse (grands pointillés).

### B-II.2.2.2 – LE FACTEUR $\sigma^B$ ET LA REPOSE GENERALE AU STRESS CHEZ *BACILLUS SUBTILIS*

Le facteur  $\sigma^B$  est impliqué dans la réponse générale au stress chez *Bacillus subtilis* (Hecker *et al.*, 1996), à l'image du facteur  $\sigma$  RpoS d'*E. coli*.

L'expression du gène *sigB* est activée par de nombreuses conditions de stress (choc thermique, alcool, acide, stress osmotique, épuisement en phosphate, glucose ou oxygène) mais aussi par l'entrée de *B. subtilis* en phase stationnaire sur milieu riche.

Le gène *sigB* fait partie d'une structure opéronique incluant 7 autres gènes (*rsbR*, *rsbS*, *rsbT*, *rsbU*, *rsbV*, *rsbW*, *sigB*, *rsbX*) et est transcrit à partir de deux promoteurs. Le premier, reconnu par le facteur  $\sigma$  principal  $\sigma^A$ , entraîne l'expression basale des 8 gènes durant la croissance (Wise and Price, 1995). Le second, localisé en amont de *rsbV*, est régulé par le facteur  $\sigma^B$  lui-même et entraîne donc l'induction de l'expression des gènes *rsbV*, *rsbW*, *sigB* et *rsbX* dans les conditions où le facteur  $\sigma^B$  est activé (pour revue, (Haldenwang, 1995)).

Un système complexe de transduction du signal impliquant notamment un système anti-anti- $\sigma$  / anti- $\sigma$  est à l'origine de l'induction de la transcription des gènes de réponse aux stress par le facteur  $\sigma^B$  (Figure I-23, pour revue (Hecker and Volker, 1998)).

L'activité du facteur  $\sigma^B$  est en effet contrôlée par un anti-facteur  $\sigma$ , RsbW, qui fixe et donc inhibe le facteur  $\sigma^B$  durant la phase exponentielle. En réponse aux stress ou à la carence nutritionnelle, RsbW est alors « capturé » par l'anti-anti-facteur  $\sigma$  RsbV, libérant alors le facteur  $\sigma^B$  actif. L'activité du facteur RsbV est elle-même régulée par un système complexe de phospho-relais, impliquant notamment l'activité kinase de RsbW. Une carence en ATP, résultant de l'épuisement du milieu, inhibe la phosphorylation de RsbV par RsbW. Les signaux de stress sont quant à eux perçus par l'intermédiaire d'un autre phospho-relai, stimulant l'activité de la phosphatase RsbU, spécifique de l'anti-anti-facteur  $\sigma$  RsbV. Ainsi, les signaux de stress et de carence nutritionnelle sont tous deux intégrés en augmentant la quantité de protéines RsbV déphosphorylées, capables de libérer le facteur  $\sigma^B$  de son anti-facteur  $\sigma$  RsbW. Le facteur  $\sigma^B$  peut alors interagir avec le core enzyme de l'ARN polymérase et ainsi induire la transcription de ces gènes cibles et donc de son propre gène.

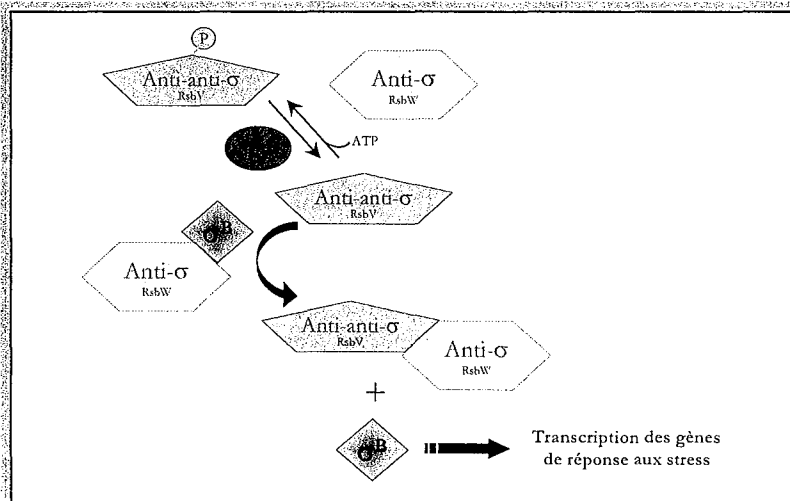


Figure I-23 : Modèle de régulation du facteur  $\sigma^B$  par un système anti- $\sigma$  / anti-anti- $\sigma$ .

Le facteur  $\sigma^B$  contrôle un grand nombre de gènes codant des protéines impliquées dans la réponse générale au stress (protéines GSP). En revanche, il ne serait pas responsable de l'induction de l'expression des gènes codant les protéines spécifiques du choc thermique (protéines HSP) (Volker *et al.*, 1994).

Différentes approches visant à identifier les gènes cibles du facteur  $\sigma^B$  ont permis d'estimer le nombre de ces cibles à plus d'une centaine (Hecker and Volker, 1998; Petersohn *et al.*, 2001). Cependant peu d'études relatent la fonction des protéines induites par le facteur  $\sigma^B$ . Les gènes codant les protéases ATP-dépendantes Clp appartiendraient à ce régulon. Ces protéases, nécessaires à la réponse générale au stress, permettraient de détoxiquer la cellule des protéines dénaturées suite à un stress (Gottesman, 1996; Gerth *et al.*, 1998).

Parmi ces gènes, il est également intéressant de noter la présence des gènes *katE*, (codant une catalase) et *dps* (codant des protéines de fixation à l'ADN), dont les homologues chez *E. coli*, sont eux aussi régulés par le facteur  $\sigma$  de réponse générale au stress RpoS (Volker *et al.*, 1994; Engelmann *et al.*, 1995).

### **B-II.2.3 – LES FACTEURS SIGMA ALTERNATIFS CHEZ STREPTOMYCES**

#### B-II.2.3.1 – LA DIFFERENCIATION ET LA REPOSE AUX STRESS CHEZ STREPTOMYCES COELICOLOR

##### B-II.2.3.1.1 - Les facteurs $\sigma$ WhiG et $\sigma^F$ chez *Streptomyces coelicolor*

Le facteur  $\sigma$  WhiG intervient lors de la mise en place du processus de sporulation. En effet, l'interruption de *whiG* n'empêche pas la croissance végétative de la bactérie mais les hyphes aériens ne se différencient plus (Chater *et al.*, 1989). En contrepartie, la surexpression de *whiG* induit une sporulation accrue dans le mycélium aérien ainsi qu'une sporulation ectopique abondante dans le mycélium végétatif. Hiérarchiquement, *whiG* se situe en amont de tous les autres loci *whi* testés et de *sigF* (Kelemen *et al.*, 1996).

Le facteur  $\sigma$  WhiG, acteur de la différenciation tardive chez les *Streptomyces*, présente une transcription constitutive à partir d'un site unique d'initiation de la transcription (Kelemen *et al.*, 1996). Impliqué dans la formation des spores,  $\sigma$  WhiG présente une activité durant les phases tardives. Des données suggèrent fortement que l'activité de  $\sigma$  WhiG serait régulée par un anti- $\sigma$ . En effet,  $\sigma$  WhiG complète l'action de son  $\sigma$  homologue chez *S. typhimurium*,  $\sigma^{FliA}$ , impliqué dans la synthèse des flagelles. Un contrôle négatif de l'activité de  $\sigma^{WhiG}$  a été mis en évidence par l'action du produit du gène *flgM* codant l'anti- $\sigma^{FliA}$  (Kelemen *et al.*, 1996).



L'inactivation du gène *sigF* entraîne la formation de spores dont la paroi est anormalement fine et les colonies mutées ont un phénotype "white". Le facteur  $\sigma^F$  est donc impliqué dans les étapes tardives de la maturation des spores (Potuckova *et al.*, 1995; Rezuchova *et al.*, 1997).

La mutation nulle de *whiG* est épistatique sur l'expression de *sigF*. Ainsi, une hiérarchie entre *whiG* et *sigF* a été dégagée avec WhiG intervenant en amont de SigF dans la cascade de facteurs  $\sigma$ . Cependant, aucune séquence régulatrice spécifique de  $\sigma$  WhiG n'a pu être détectée dans les régions promotrices du gène *sigF*, ce qui sous-entend donc l'existence d'au moins un autre facteur  $\sigma$  alternatif (Kelemen *et al.*, 1996).

#### **B-II.2.3.1.2 - Le facteur $\sigma$ BldN chez *Streptomyces coelicolor***

Le gène *bldN*, préalablement nommé *whiN*, code un facteur  $\sigma$  de type ECF nécessaire à la formation du mycélium aérien chez *S. coelicolor* (Bibb *et al.*, 2000).

Le promoteur du gène *bldM*, un autre locus impliqué dans le développement de *S. coelicolor*, est reconnu par le facteur  $\sigma$  BldN. Ce facteur  $\sigma$  présente une extension N-terminale suggérant une régulation post-traductionnelle de type maturation de pro- $\sigma$ .

#### **B-II.2.3.1.3 - Le facteur $\sigma^H$ chez *Streptomyces coelicolor***

Plus récemment, un troisième facteur  $\sigma$  codé par le gène *sigH* a été identifié. Un mutant *sigH* présente, à maturité, une coloration gris-pâle. La formation du mycélium aérien n'est pas affectée mais la différenciation des spores est incomplète. Les rares spores formées ne montrent toutefois pas de différence morphologique par rapport aux spores de la souche sauvage. De plus, ce facteur  $\sigma^H$  est impliqué dans la réponse au stress osmotique (Sevcikova *et al.*, 2001). Ce facteur  $\sigma^H$  est donc au carrefour du stress et du développement chez les *Streptomyces* (Kelemen *et al.*, 2001).

La régulation de ce facteur  $\sigma$  se fait à de multiples niveaux (voir Discussion) : au niveau transcriptionnel, via la présence de promoteurs multiples en amont du gène *sigH* (Kormanec *et al.*, 2000; Kelemen *et al.*, 2001), au niveau traductionnel via l'utilisation de codons d'initiation alternatifs et également au niveau post-traductionnel par des mécanismes de maturation de pro- $\sigma$  et des interactions avec un anti-facteur  $\sigma$  RsrA (Sevcikova and Kormanec, 2002; Viollier *et al.*, 2003b). Deux gènes cibles du facteur  $\sigma^H$  ont été identifiés : *srgB*, codant un homologue d'une protéine impliquée dans la sporulation et la division cellulaire (Kormanec and Sevcikova, 2002a) et *gltB* codant une glutamate synthétase (Kormanec and Sevcikova, 2002b).

Le facteur  $\sigma^H$  appartiendrait à une famille de 9 facteurs  $\sigma$  chez *S. coelicolor*, homologues du facteur  $\sigma^B$  de réponse générale au stress de *B. subtilis* (Viollier *et al.*, 2003a).

**B-II.2.3.1.4 - Le facteur  $\sigma^B$  chez *Streptomyces coelicolor***

Le gène, *sigB*, codant un facteur  $\sigma$  alternatif impliqué dans la réponse au stress osmotique chez *S. coelicolor* A3(2), fait également partie de cette famille de facteurs  $\sigma$ , homologues du facteur  $\sigma^B$  de *B. subtilis*. La mutation de *sigB* chez *S. coelicolor* A3(2) conduit au phénotype « bald » des colonies sur un milieu spécifique (Cho *et al.*, 2001).

Les facteurs  $\sigma^H$  et  $\sigma^B$  sont impliqués à la fois dans la différenciation et la réponse au stress. Des études transcriptionnelles des gènes *sigH* et *sigB* montrent que leur expression est induite en réponse à un stress osmotique. La croissance des mutants *sigH* et *sigB* est affectée lors d'un stress osmotique par rapport à celle d'une souche non mutée (Cho *et al.*, 2001).

Ces données suggèrent l'existence d'une connexion intime entre la différenciation morphologique et le stress chez *S. coelicolor* A3(2). La pression de sélection aurait conduit la bactérie à détourner des systèmes impliqués dans la réponse au stress vers une partie du programme de son développement.

**B-II.2.3.1.5 - Le facteur  $\sigma^R$  chez *Streptomyces coelicolor***

Le facteur sigma  $\sigma^R$  est un facteur  $\sigma$  de type ECF, impliqué dans la réponse au stress oxydatif. Sa délétion provoque une sensibilité aux composés générateurs de radicaux superoxydes ainsi qu'à la diamide (un agent oxydant les résidus thiols) mais pas à l' $H_2O_2$  (Paget *et al.*, 1998). Il contrôle des gènes de réponse typique au stress oxydatif, tel que l'opéron *trxB* codant le système thiorédoxine/thiorédoxine réductase (Paget *et al.*, 1998), mais également des gènes codant des sous-unités du ribosome ou de l'ARN polymérase, suggérant que les machineries de transcription et de traduction sont modifiées suite à un stress disulfide (Paget *et al.*, 2001b).

La cascade de signalisation permettant d'aboutir à la surexpression des gènes de réponse au stress oxydatif implique l'intervention d'un anti-facteur  $\sigma$ , RsrA, lui même sensible à l'oxydation (Kang *et al.*, 1999), voir paragraphe A-1.2.3 de la discussion). Ce facteur  $\sigma^R$  est également régulé au niveau traductionnel, puisque synthétisé sous forme d'un pro- $\sigma^R$  (voir paragraphe A-1.2.1 de la discussion).

**B-II.2.3.2.3 - Le facteur  $\sigma$  sigG chez *Streptomyces coelicolor***

Le facteur  $\sigma$  SigG a été caractérisé suite au criblage d'une banque à l'aide d'une sonde correspondant à un motif conservé de la région 2.2 des facteurs  $\sigma$  (Kormanec *et al.*, 1999). Ce facteur  $\sigma$  est fortement homologue du facteur  $\sigma^F$  de *S. coelicolor* et du facteur de réponse générale au stress  $\sigma^B$  de *B. subtilis*, notamment dans les régions 2.4 et 4.2, impliquées dans la reconnaissance des promoteurs. Cependant aucune donnée n'existe concernant son expression ni la réponse dans laquelle il serait impliqué.

B-II.2.3.3 – LES (AUTRES) FACTEURS  $\sigma$  DE TYPE ECF CHEZ  
STREPTOMYCES COELICOLOR

B-II.2.3.3.1 - Le facteur  $\sigma^E$  chez *Streptomyces coelicolor*

Le facteur  $\sigma$  alternatif extracellulaire  $\sigma^E$  est impliqué dans la formation des parois cellulaires chez *S. coelicolor* (Paget *et al.*, 1999). Le seul gène connu pour être sous contrôle du facteur  $\sigma^E$  est le gène *brdD*, codant le facteur  $\sigma^{\text{HrdD}}$ , un des trois facteurs  $\sigma$  de *S. coelicolor* de fonction inconnue homologues du facteur sigma principal  $\sigma^{\text{HrdB}}$  (Buttner *et al.*, 1990).

Le promoteur du gène *sigE* est similaire à celui du gène *whiG*, codant le facteur  $\sigma$  spécifique de la sporulation chez *S. coelicolor*, mais le facteur  $\sigma$  responsable de la reconnaissance de ce type de promoteur reste inconnu (Paget *et al.*, 1999).

Un modèle de régulation de l'activité de  $\sigma^E$  en réponse à des signaux provenant de l'enveloppe cellulaire a été proposé (JHong *et al.*, 2002) (Figure I-24).

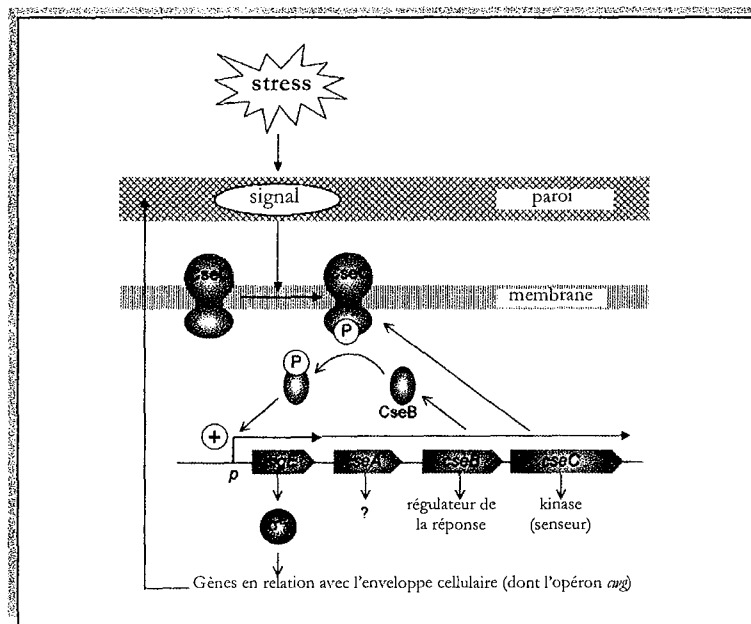


Figure I-24 : Modèle de régulation de l'activité de  $\sigma^E$  en réponse à des signaux provenant de l'enveloppe cellulaire. (d'après Hong *et al.* (2002)).

#### B-II.2.3.3.2 - Le facteur $\sigma^U$ chez *Streptomyces coelicolor*

Le gène *sigU* code un facteur  $\sigma$  de type ECF chez *S. coelicolor*. La dérégulation de ce facteur  $\sigma$ , entraînée par l'interruption du gène codant son anti-s RsuA, affecte la différenciation à un stade précoce, mais aucun phénotype n'est associé à la mutation du facteur  $\sigma$  (Gehring *et al.*, 2001).

#### B-II.2.3.4 - CHEZ LES AUTRES STREPTOMYCES : ILLUSTRATION DE LA DIVERSITE DES FACTEURS SIGMA

L'expression du métabolisme secondaire est associée de façon étroite aux étapes tardives de différenciation.

Un facteur  $\sigma$  alternatif appartenant à la famille des ECF, codé par le gène *sigE*, a été directement impliqué dans la régulation de la production d'actinomycine chez *S. antibioticus* (Jones *et al.*, 1997).

Un facteur  $\sigma$ , homologue du facteur  $\sigma^B$  de *B. subtilis*, a été impliqué dans la biosynthèse des caroténoïdes chez *S. setonii* (Kato *et al.*, 1995) et *S. griseus* (Lee *et al.*, 2001).

∞ MATERIEL

ET

METHODES ∞

- CHAPITRE I -  
MANIPULATION DES SOUCHES BACTERIENNES ET VECTEURS DE CLONAGE

<b>I.1</b>	<b>œ Souches bactériennes, vecteurs et conditions de croissance</b>	<b>75</b>
I.1.1	- Souches bactériennes	75
I.1.2	- Vecteurs de clonage	76
I.1.3	- Milieux et conditions de culture	79
I.1.4	- Courbes de croissance	81
I.1.4.1	- Courbes de croissance par mesure de la concentration en ADN	81
I.1.4.2	- Courbes de croissance par mesure de la densité optique	83
<b>I.2</b>	<b>œ Transformation de souches bactériennes</b>	<b>84</b>
I.2.1	- Electroporation d' <i>E. coli</i>	84
I.2.1.1	- Introduction des cosmides dans <i>E. coli</i> BW25113/pIJ790	84
I.2.1.2	- Construction des cosmides modifiés	85
I.2.2	- Conjugaison intraspécifique <i>E. coli</i> S17 / <i>E. coli</i> SURE™.	85
I.2.3	- Conjugaison intergénérique <i>E. coli</i> / <i>S. ambofaciens</i>	85
I.2.4	- Transformation de protoplastes de <i>Streptomyces</i>	87
<b>I.3</b>	<b>œ Observation des colonies bactériennes à la recherche d'un phénotype</b>	<b>88</b>
I.3.1	- Observations morphologiques	88
I.3.2	- Traitements aux agents oxydants chimiques	88
I.3.3	- Traitement aux rayonnements ultraviolets (UV)	88
I.3.4	- Détermination de l'activité antibactérienne	89

- CHAPITRE II -  
MANIPULATION DES ACIDES NUCLEIQUES

<b>II.1</b>	<b>œ Préparation des acides nucléiques</b>	<b>91</b>
II.1.1	- Préparation d'ADN, digestion enzymatique et électrophorèse	91
II.1.1.1	- Extraction et purification d'ADN	91
II.1.1.2	- Digestions par les endonucléases et électrophorèse d'ADN	91
II.1.1.3	- Ligation	92
II.1.2	- Préparation et dosage d'ARN	93

<b>II.2 ☞ Les techniques utilisant les acides nucléiques</b>	<b>94</b>
II.2.1 - Hybridation ADN - ADN	94
II.2.2 - Séquençage automatique d'ADN	94
II.2.3 - Transcription inverse (RT) et amplification d'ADN (PCR)	94
II.2.3.1 - Difficultés de la méthode dans le cas des bactéries	94
II.2.3.2 - Choix d'un témoin interne pour une approche qualitative : le gène <i>recA</i>	96
II.2.3.3 - Les milieux réactionnels de RT-PCR	97
II.2.3.3.1 - La transcription inverse des ARN en ADNc	97
II.2.3.3.2 - L'amplification d'ADN par polymérisation en chaîne (PCR)	98
II.2.4 - Extension d'amorces	101
II.2.5 - Cartographie à la nucléase S1	102
<b>II.3 ☞ Analyse des séquences d'ADN</b>	<b>103</b>
II.3.1 - Détection de phases ouvertes de lectures	103
II.3.2 - Analyse de la redondance d'un génome par le logiciel BLASTCLUST	103

**- CHAPITRE III -  
LA PCR QUANTITATIVE**

<b>III.1 ☞ Principe de la PCR quantitative</b>	<b>105</b>
III.1.1 - Les bases de la PCR quantitative en temps réel	105
III.1.2 - La technologie de détection basée sur le SYBR Green	105
III.1.3 - La notion de cycle seuil ou Ct (threshold cycle)	107
III.1.4 - Efficacité et spécificité de la réaction	108
III.1.4.1 - Efficacité de la PCR et courbe standard	108
III.1.4.2 - Spécificité de la PCR et courbe de fusion	109
<b>III.2 ☞ Un exemple d'analyse de résultats de PCR quantitative</b>	<b>111</b>

- CHAPITRE I -

MANIPULATION DES SOUCHES  
BACTERIENNES ET VECTEURS DE  
CLONAGE



## I.1 $\otimes$ SOUCHES BACTERIENNES, VECTEURS ET CONDITIONS DE CROISSANCE

### I.1.1 - SOUCHES BACTERIENNES

Les souches sauvage et mutantes de *S. ambofaciens* DSM40697, les souches *E. coli* BW25113, ET12567 et DH5 $\alpha$  ainsi que la souche de *B. subtilis* ATCC6633 utilisées dans ce travail sont décrites dans le Tableau MM-1. *S. lividans* ZX7 sert d'hôte pour le clonage de fragments d'ADN et les expériences impliquant un gène rapporteur (Zhou *et al.*, 1988). Les souches de *S. lividans* transformées par les constructions plasmidiques réalisées au cours de ces travaux seront désignées NSL pour Nancy *Streptomyces lividans*. Les souches mutantes *S. ambofaciens* NSA855 et NSAH135 sont issues de l'instabilité génétique (Leblond *et al.*, 1991; Fischer *et al.*, 1998b; Wenner, 1998). La souche *Escherichia coli* JM101 a été utilisée comme hôte pour la propagation des dérivés recombinants du bactériophage M13. Son génotype (Stratagene) est : *supE thi*  $\Delta$ (*lac-proAB*) F' [*traD proAB<sup>+</sup> lacI<sup>f</sup> lacZ* $\Delta$ M15].

Souches	Génotype	Référence
<b><i>Streptomyces ambofaciens</i></b>		
DSM40697		DSM *
$\Delta$ <i>hasL1</i>	$\Delta$ <i>hasL::Apra</i>	Ce travail
$\Delta$ <i>hasL2</i>	$\Delta$ <i>hasL::Apra</i>	Ce travail
$\Delta$ <i>hasL3</i>	$\Delta$ <i>hasL::Apra</i>	Ce travail
$\Delta$ <i>hasL4</i>	$\Delta$ <i>hasL::Apra</i>	Ce travail
$\Delta$ <i>hasR1</i>	$\Delta$ <i>hasR::Apra</i>	Ce travail
$\Delta$ <i>hasR2</i>	$\Delta$ <i>hasR::Apra</i>	Ce travail
$\Delta$ <i>hasR3</i>	$\Delta$ <i>hasR::Apra</i>	Ce travail
$\Delta$ <i>hasR4</i>	$\Delta$ <i>hasR::Apra</i>	Ce travail
$\Delta$ <i>hasL</i> $\Delta$ <i>hasR</i> DMLR1	$\Delta$ <i>hasL::Apra</i> , $\Delta$ <i>hasR::Spec</i>	Ce travail
$\Delta$ <i>hasL</i> $\Delta$ <i>hasR</i> DMLR2	$\Delta$ <i>hasL::Apra</i> , $\Delta$ <i>hasR::Spec</i>	Ce travail
$\Delta$ <i>hasL</i> $\Delta$ <i>hasR</i> DMLR3	$\Delta$ <i>hasL::Apra</i> , $\Delta$ <i>hasR::Spec</i>	Ce travail
$\Delta$ <i>hasR</i> $\Delta$ <i>hasL</i> DMRL1	$\Delta$ <i>hasR::Apra</i> , $\Delta$ <i>hasL::Spec</i>	Ce travail
$\Delta$ <i>hasR</i> $\Delta$ <i>hasL</i> DMRL2	$\Delta$ <i>hasR::Apra</i> , $\Delta$ <i>hasL::Spec</i>	Ce travail
$\Delta$ <i>hasR</i> $\Delta$ <i>hasL</i> DMRL3	$\Delta$ <i>hasR::Apra</i> , $\Delta$ <i>hasL::Spec</i>	Ce travail
<b><i>Escherichia coli</i></b>		
BW25113	$\Delta$ ( <i>araD-araB</i> )567 $\Delta$ <i>lacZ</i> 4787(:: <i>rrnB-A</i> ) <i>lacI</i> p-4000( <i>lacI<sup>Q</sup></i> ) $\lambda$ ;	(Datsenko and
	<i>rpoS</i> 369(Am) <i>rph-1</i> $\Delta$ ( <i>rbaD-rbaB</i> )568 <i>hsdR</i> 514	Wanner, 2000)
ET12567	<i>dam dcm hsdS</i> Cm <sup>R</sup>	(MacNeil, 1988)
DH5 $\alpha$	<i>supE</i> 44 $\Delta$ <i>lacU</i> 169 ( $\Phi$ 80 <i>lacZ</i> $\Delta$ M15) <i>hsdR</i> 17 <i>recA</i> 1 <i>endA</i> 1	(Hanahan, 1983)
<b><i>Bacillus subtilis</i></b>		
ATCC6633		ATCC **

Tableau MM-1 : Souches utilisées lors de ce travail.

Les simples mutants sont issus de la souche *S. ambofaciens* DSM40697. Les doubles mutants DMLR et DMRL sont issus respectivement des souches  $\Delta$ *hasL3* et  $\Delta$ *hasR2*. \*DSM : Deutsche Sammlung von Mikroorganismen. \*\*ATCC : American Type Culture Collection.

### I.1.2 - VECTEURS DE CLONAGE

Le plasmide pIJ487 (Ward *et al.*, 1986), utilisé lors des expériences de clonage des promoteurs des gènes *has* en amont du gène rapporteur *neo*, est un dérivé de pIJ101 (Figure MM-1). Il s'agit d'un plasmide à grand nombre de copies portant un gène de résistance au thiostrepton. Il présente en aval d'un site de clonage multiple le gène rapporteur *neo* codant une aminoglycoside phosphotransférase conférant la résistance dose-dépendante à la kanamycine. En amont du site de clonage multiple se trouve un terminateur de transcription. La partie codante du gène *neo* est précédée d'un RBS contenant un codon d'arrêt de traduction. Dans notre cas, le codon "start" de l'ORF *has* est alors en phase avec le codon "stop" situé dans le RBS du gène *neo*. Ainsi, la traduction initiée au codon d'initiation des ORF *has* est arrêtée. Il n'y aura pas de fusion traductionnelle mais seule la traduction du gène *neo* sera décelable, si un promoteur de transcription est inséré en amont de celui-ci.

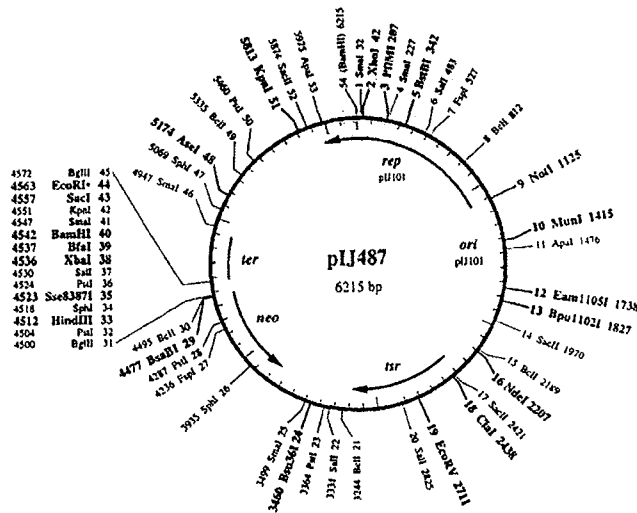


Figure MM-1 : Carte du plasmide pIJ487

Le plasmide conjugatif et intégratif pIJ8630 (Figure MM-2) a été utilisé pour le clonage de la région promotrice des gènes *has* en amont du gène rapporteur codant l'EGFP et de fait, lors des expériences de conjugaison entre *E.coli* S17-1 et *S.ambifaciens* DSM40697 (Sun *et al.*, 1999). Le vecteur pIJ8630 a été construit à partir du plasmide pIJ8600, lui-même dérivé du plasmide conjugatif et intégratif pSET152 (Bierman *et al.*, 1992).

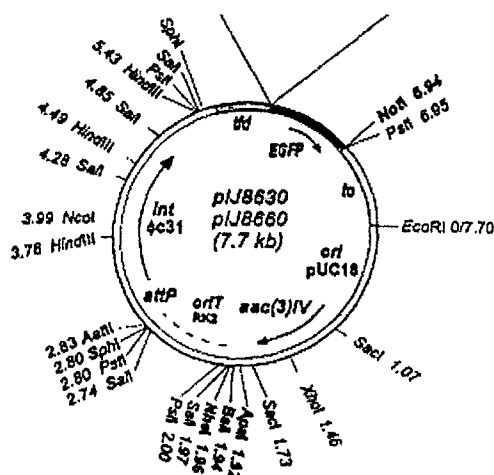


Figure MM-2 : Carte du plasmide pIJ8630

Le plasmide pIJ8630 contient les différents éléments nécessaires à la conjugaison bactérienne, à l'intégration sur le chromosome de *Streptomyces* et à l'étude de l'expression d'un gène par la méthode du gène rapporteur, soient :

- ❑ une origine de réplication (pUC18) indispensable à sa propagation chez *E. coli*,
- ❑ un gène de sélection et de maintien dans les cellules donneuses et receveuses (résistance à l'apramycine),
- ❑ une origine de transfert (*oriT* RK2),
- ❑ une région ( $\Phi$ C31 *attP*-intégrase) permettant l'intégration stable au site *attB* dans le chromosome de *Streptomyces*,
- ❑ l'ORF codant la protéine EGFP (Enhanced Green Fluorescent Protein) flanquée de terminateurs afin d'éviter tout départ de transcription incongru. Cette ORF est immédiatement précédée d'une région contenant des sites de restriction uniques (MCS pour Multiple Cloning Site) permettant l'insertion d'une séquence d'ADN particulière.
- ❑ Des codons stop dans les trois phases de lecture entre la région MCS et le codon start de l'EGFP et ce, afin de prévenir la formation de fusions traductionnelles.

Les plasmides pIJ790, pIJ773, pIJ778 et les cosmides 14C4, 25E1 utilisés pour les expériences d'interruption des gènes *has* sont décrits en Figures MM-3, MM-4, MM-5 et MM-6. Les plasmides pIJ773 et pIJ778 portent les cassettes utilisées pour les remplacements de gènes. Ces cassettes se composent, pour pIJ773, du gène *aac(3)IV* conférant la résistance à l'apramycine et de l'origine de transfert *oriT* du plasmide RK2 (Pansegrau *et al.*, 1994), et pour pIJ778, du gène *aadA* conférant la résistance aux antibiotiques spectinomycine et streptomycine, ainsi que du locus *oriT*. Chacune des cassettes est flanquée des sites FRT pour "FLP Recognition Target". Par la suite les cassettes sont appelées "cassette Apra" (résistance à l'apramycine) et "cassette Spec" (résistance à la spectinomycine/streptomycine).

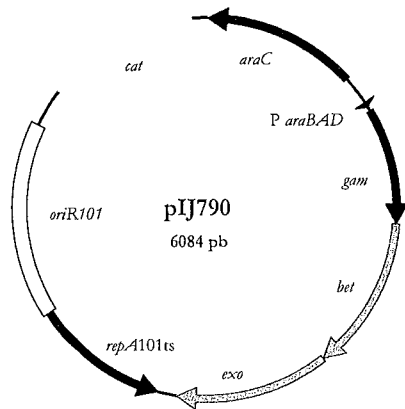


Figure MM-3 : Carte du plasmide pIJ790 (d'après Gust *et al.*, 2002).

Le plasmide pIJ790 dérive du plasmide pKD20 (Datsenko and Wanner, 2000) portant le système de recombinaison IRED. Le gène *bla* de pKD20 conférant la résistance à l'ampicilline a été remplacé dans pIJ790 par le gène *cat* conférant la résistance au chloramphénicol.

*araC* : gène codant l'activateur transcriptionnel AraC. *bet* : gène codant la protéine Bet, une protéine se fixant à l'ADN simple brin. *cat* : gène codant une chloramphénicol acétyl transférase. *exo* : gène codant l'exonucléase qui promeut la recombinaison avec Bet. *gam* : gène codant la protéine Gam, un inhibiteur de l'exonucléase V de l'hôte. *oriR101* : origine de réplication. *P araBAD*: promoteur de l'opéron arabinose inducible par le L-arabinose. *repA101ts*: origine de réplication thermosensible.

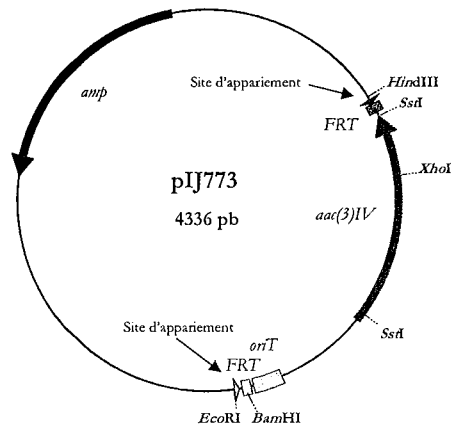


Figure MM-4 : Carte du plasmide pIJ773 (d'après Gust *et al.*, 2002).

Le plasmide pIJ773 porte la cassette de résistance à l'apramycine utilisée pour le remplacement de gène par la technique de mutagenèse ciblée par PCR.

*amp* : gène codant une b-lactamase. Confère la résistance à l'ampicilline.

*aac(3)IV* : gène codant l'apramycine acétyltransférase. Confère la résistance à l'apramycine.

*oriT* : origine de transfert du plasmide RK2.

FRT : (pour "FLP Recognition Targets") sites reconnus par la FLP recombinase.

"Site d'appariement" : sites d'appariement des amorces utilisées pour amplifier par PCR la cassette de résistance.

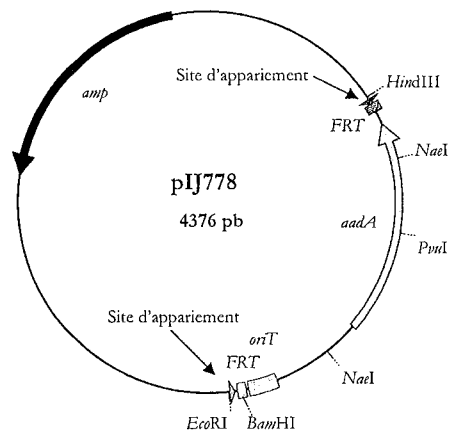


Figure MM-5 : Carte du plasmide pIJ778 (d'après Gust *et al.*, 2002).

Le plasmide pIJ778 porte la cassette de résistance à la spectinomycine/streptomycine utilisée pour le remplacement de gène par la technique de mutagenèse ciblée par PCR.

*amp* : gène codant une b-lactamase. Confère la résistance à l'ampicilline.

*aadA* : gène codant la spectinomycine adényltransférase. Confère la résistance à la spectinomycine et à la streptomycine.

*oriT* : origine de transfert du plasmide RK2.

FRT : (pour "FLP Recognition Targets") sites reconnus par la FLP recombinase.

"Site d'appariement" : sites d'appariement des amorces utilisées pour amplifier par PCR la cassette de résistance.



<b>1/1/1</b>	
L-asparagine	0,5 g
K <sub>2</sub> HPO <sub>4</sub>	0,5 g
MgSO <sub>4</sub> , 7H <sub>2</sub> O	0,2 g
FeSO <sub>4</sub> , 7H <sub>2</sub> O	0,01 g
Agar	10 g
H <sub>2</sub> O UP	qsp 1 l

pH ajusté à : 7,2

<b>Pour MMG :</b>	
Glucose	10 g
<b>Pour MMM :</b>	
Mannitol	5 g

Nutrient Agar	23 g
H <sub>2</sub> O UP	qsp 1 l

Extrait de levure	1 g
Extrait de viande	1 g
NZamine A	2 g
Dextrine	10 g
CoCl <sub>2</sub> , 6H <sub>2</sub> O	20 mg
Agar	20 g
H <sub>2</sub> O UP	qsp 1 l

pH ajusté à : 7,3

Saccharose	103 g
Glucose	10 g
K <sub>2</sub> SO <sub>4</sub>	0,25 g
MgCl <sub>2</sub>	10,12 g
Casaminoacides	0,1 g
H <sub>2</sub> O UP	qsp 800 ml
Puis ajout de :	
KH <sub>2</sub> PO <sub>4</sub> (0,5 %)	10 ml
CaCl <sub>2</sub> , 2H <sub>2</sub> O (3,68 %)	80 ml
L-proline (20 %)	15 ml
Tampon TES (5,73 %)	100 ml
*Solution d'éléments	2 ml
traces	
NaOH 1M	5 ml

\*Solution d'éléments

traces (l<sup>-1</sup>) :

ZnCl <sub>2</sub>	40 mg
FeCl <sub>3</sub> , 6H <sub>2</sub> O	200 mg
CuCl <sub>2</sub> , 2H <sub>2</sub> O	10 mg
MnCl <sub>2</sub> , 4H <sub>2</sub> O	10 mg
NaB <sub>4</sub> O <sub>7</sub> , 10H <sub>2</sub> O	10 mg
(NH <sub>4</sub> ) <sub>6</sub> Mo <sub>7</sub> O <sub>24</sub> , 4H <sub>2</sub> O	10 mg

**Pour R2YE :**

Extrait de levure (10 %)	5 ml
--------------------------	------

**Tableau MM-2 :** Composition des milieux de culture utilisés lors des observations morphologiques.

MM : Milieu Minimum (Hopwood, 1967). MMG : Milieu Minimum Glucose. MMM : Milieu Minimum Mannitol. HT : Hickey-Tresner (Pridham *et al.*, 1957a). NA : Nutrient Agar. R2 : (Okanishi *et al.*, 1974). R2YE : (Thompson *et al.*, 1980).

Les cultures d'*E. coli* JM101 sont réalisées en milieu liquide 2 x YT (Sambrook *et al.*, 1989) à 37°C. Lors de l'infection par les dérivés du bactériophage M13, une préculture (37°C pendant 12 h) est diluée au 100<sup>ème</sup>, puis cultivée sous agitation pendant 4 h à 350 rpm. Sur milieu solide, *E. coli* JM101 est cultivée sur milieu LB (Sambrook *et al.*, 1989). Les souches sont conservées à -20°C et -70°C dans du milieu de culture additionné de glycérol à la concentration finale de 20%.

Les *Streptomyces* sont cultivés en milieu HT, liquide et gélosé (Pridham *et al.*, 1957a). Le milieu YEME est utilisé pour les cultures de *Streptomyces* en vue de la préparation de protoplastes. Les méthodes de protoplastisation, de titration et de régénération des protoplastes sont décrites par (Hopwood *et al.*, 1985). Le milieu de régénération (milieu R5) est additionné de thiostrepton (à la concentration finale de 30 µg/ml) pour la sélection des transformants, et de kanamycine (concentration finale de 25 µg/ml à 800 µg/ml) pour l'expression de la résistance depuis le gène rapporteur *néo*. Les spores de *Streptomyces* sont conservées dans du glycérol à la concentration finale de 20% à -20°C.

## I.1.4 - COURBES DE CROISSANCE

### I.1.4.1 - COURBES DE CROISSANCE PAR MESURE DE LA CONCENTRATION EN ADN

Les courbes de croissance sont réalisées à partir de cultures de 200 ml en milieu HTensemencés à 10<sup>8</sup> UFC/ml. Des aliquots (1 ml) du milieu de culture sont prélevés à des temps définis au préalable (3 prélèvements par temps). Les échantillons sont ensuite soniqués (60 W pendant 1 min) afin de lyser le mycélium, puis l'ADN libéré est dosé suivant la méthode de Burton (Burton, 1956). En solution acide, la forme linéaire des désoxyntoses de l'ADN est convertie en β-hydroxylévalinaldéhyde, réactif qui s'associe à la diphénylamine pour former un complexe de couleur bleue présentant un maximum d'absorption à 595 nm. Les protéines réagissent également aux réactifs. La densité optique (DO) est alors mesurée à 700 nm ne révélant que la réactivité avec les protéines. Le rapport des valeurs obtenues à 595 nm et 700 nm permet, grâce à une courbe étalon réalisée avec des dilutions successives d'une solution d'ADN calibrée, de déterminer la concentration d'ADN.

Compte tenu des fonctions codées par la plupart des gènes localisés dans la région instable, c'est-à-dire impliquées dans l'expression en phase de croissance tardive, l'expression des ORF *bas* a été recherchée au cours de la croissance. La réalisation de courbes de croissance est un préalable qui permet de déterminer les temps de culture correspondant aux différentes phases de croissance. Ainsi, la croissance des différentes souches a été suivie par la mesure de la concentration en ADN au cours d'une culture en milieu HT liquide (Figures MM-7 et MM-8). Les temps de culture retenus pour les points correspondant aux phases de croissance exponentielle, de transition (entrée en phase stationnaire) et stationnaire sont respectivement 9 h, 16 h et 24 h. Le choix de ces temps a été validé en réalisant les courbes de croissance à trois reprises.

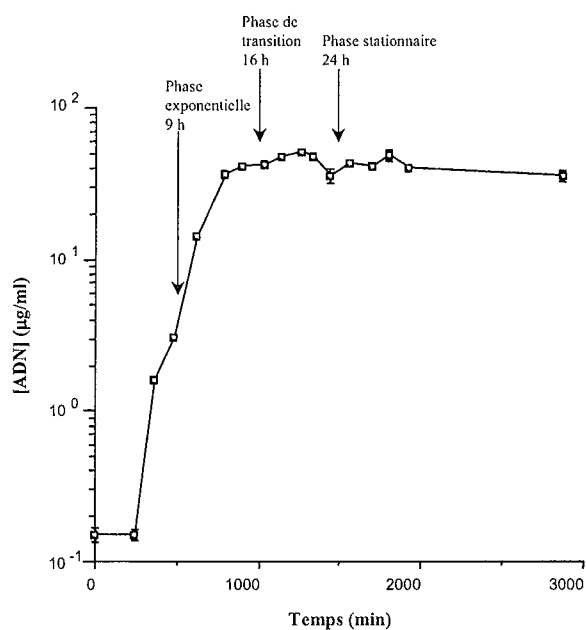


Figure MM-7 : Courbe de croissance de la souche sauvage *S. ambofaciens* DSM40697. La concentration d'ADN (µg/ml) a été déterminée sur des aliquots de culture prélevés aux différents temps selon la technique de Burton (1956). Les écarts-types sont calculés à partir des données des trois courbes de croissance.

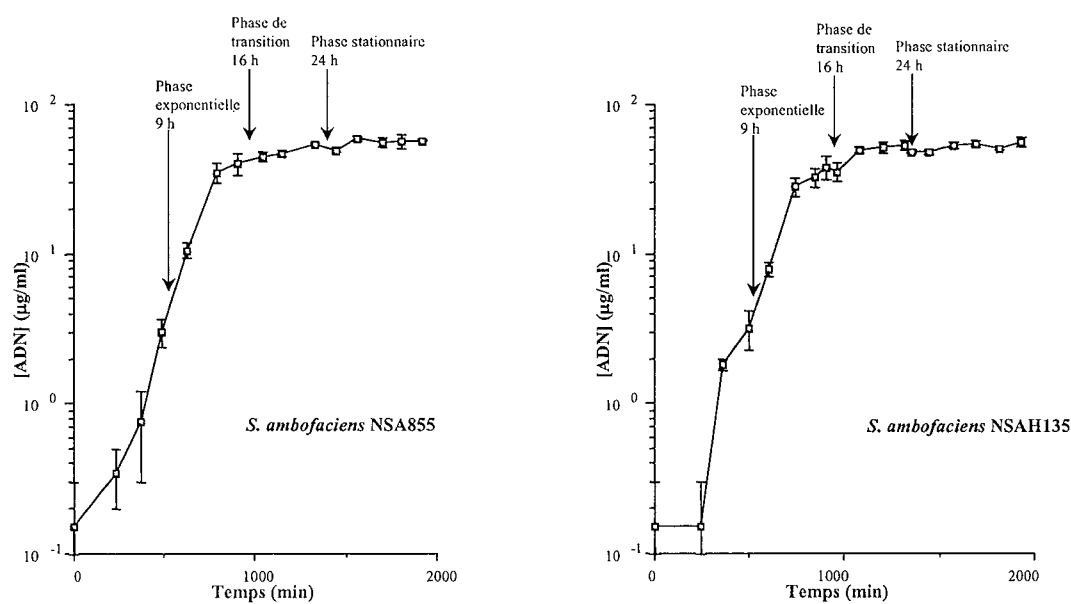


Figure MM-8 : Courbe de croissance des souches mutantes *S. ambofaciens* NSA135 et NSA855. Le symbolisme est identique à celui de la Figure MM-7.



### I.1.4.2 - COURBES DE CROISSANCE PAR MESURE DE LA DENSITÉ OPTIQUE

Lors du clonage des régions promotrices des gènes *bas* en amont du gène rapporteur *néo* conférant la résistance à la kanamycine, des courbes de croissance de la souche *S. lividans* NSL9 et de la souche témoin (*S. lividans* ZX7 transformée par pIJ487) ont été réalisées. Ainsi, une culture de chacune de ces 2 souches a été réalisée en milieu HT liquide (inoculum de  $10^6$  spores) et la croissance mesurée par le suivi de la densité optique à 660 nm (Figure MM-9).

Pour chaque temps de croissance 3 échantillons de culture sont prélevés et soniqués afin de dissocier les pelotes de mycélium. Pour chaque échantillon, la densité optique est mesurée afin d'établir une courbe de croissance (Figure MM-9).

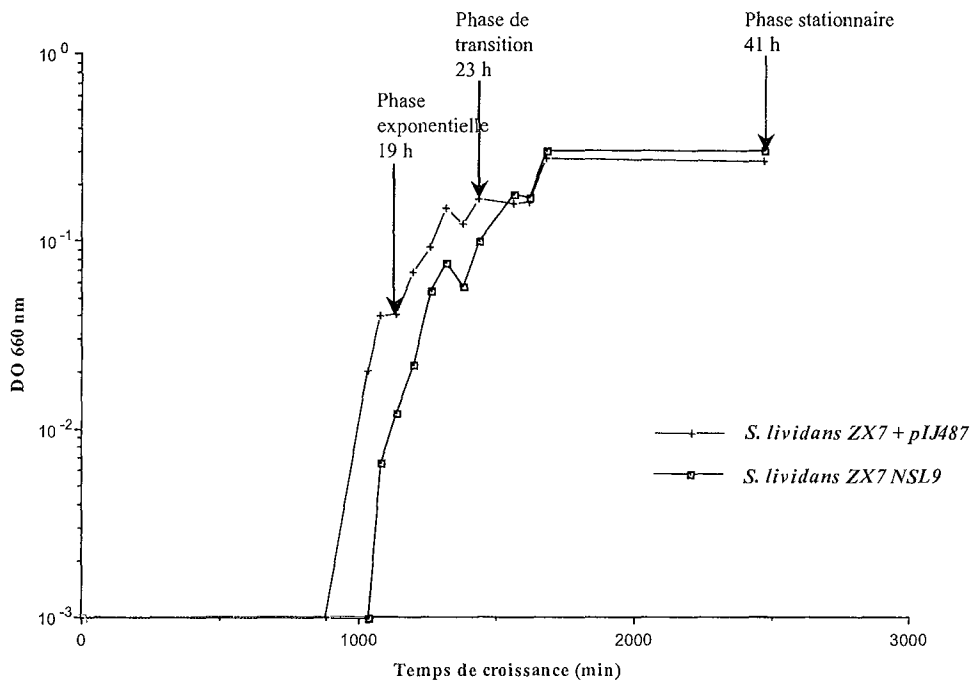


Figure MM-9 : Courbes de croissance de la souche *S. lividans* NSL9 et de la souche témoin transformée avec le plasmide natif pIJ487.

## I.2 $\infty$ TRANSFORMATION DE SOUCHES BACTERIENNES

### I.2.1 - ELECTROPORATION D'*E. COLI*

Les électroporations sont réalisées avec l'appareil GenePulserII (Biorad) calibré à 200  $\Omega$ , 25 $\mu$ F et 2,5 kV. Les constantes de temps attendues varient entre 3,8 et 4,9 ms.

Le taux de survie des cellules ayant subi un choc électrique est déterminé par le rapport entre le nombre de colonies obtenues, compte tenu des facteurs de dilution, avant et après électroporation avec 1  $\mu$ l d'eau ultra-pure.

De même, une efficacité de transformation peut être estimée en rapportant le nombre de transformants obtenus à la quantité d'ADN utilisée pour la transformation. Pour réaliser ce témoin, 1  $\mu$ l d'une dilution de l'ADN du vecteur de clonage natif (à une concentration connue) est utilisée.

Seules les transformations d'*E. coli* réalisées lors de la construction des mutants interrompus pour le(s) gène(s) *has* vous sont présentées ci-dessous.

#### I.2.1.1 - INTRODUCTION DES COSMIDES DANS *E. COLI* BW25113/pIJ790

Une préculture de *E. coli* BW25113/pIJ790 est réalisée dans 10 ml de LB contenant 25  $\mu$ g/ml de chloramphénicol sur la nuit à 30°C sous agitation à 250 rpm. Une culture est réalisée en inoculant 50 ml de LB supplémenté à 20 mM de MgSO<sub>4</sub> et 25  $\mu$ g/ml de chloramphénicol avec la préculture (1 % final). Les cellules sont mises à croître à 30°C sous agitation à 250 rpm jusqu'à obtention d'une DO<sub>600</sub> d'environ 0,6. Les cellules sont récupérées par centrifugation à 4000 rpm à 4°C pendant 5 min, lavées deux fois avec du glycérol 10 % froid puis reprises dans 100  $\mu$ l de glycérol 10 % froid. Un volume de 50  $\mu$ l de cellules compétentes est électroporé avec 100 ng de cosmide (14C4 ou 25E1). Les cellules sont immédiatement diluées dans 1 ml de LB froid puis mises à incuber à 30°C sous agitation pendant 1 h. Les cellules sont ensuite étalées sur milieu LB additionné d'ampicilline (50  $\mu$ g/ml), de kanamycine (50  $\mu$ g/ml), et de chloramphénicol (25 $\mu$ g/ml) et incubées à 30°C (température permissive pour la réplication de pIJ790) sur la nuit. Quelques clones résistants sont contrôlés pour la présence des éléments extrachromosomiques (pIJ790 et cosmide) par restriction enzymatique de leur ADN.

### I.2.1.2 - CONSTRUCTION DES COSMIDES MODIFIES

Les souches *E. coli* BW25113/pIJ790 contenant les cosmides sont transformées par électroporation (protocole analogue à celui décrit précédemment) avec 100 ng de produit de PCR correspondant aux cassettes de sélection. Seuls les cribles de sélection varient en fonction de la cassette utilisée (25 µg/ml d'apramycine pour la cassette Apra, 25 µg/ml de spectinomycine pour la cassette Spec). Afin d'induire le système λ RED, du L-arabinose est ajouté à la culture initiale (10mM final). Suite à l'électroporation, les cellules sont incubées à 37°C (température restrictive pour la réplication de pIJ790) afin de provoquer la perte du plasmide pIJ790. L'ADN cosmidique des clones positifs est vérifié par restriction enzymatique et PCR.

### **I.2.2 - CONJUGAISON INTRASPECIFIQUE *E. COLI* S17 / *E. COLI* SURE™.**

Lors des expériences de transformation de *Streptomyces* par conjugaison intergénérique, des contrôles préalables, consistant à tester la capacité de transfert des plasmides recombinants entre deux souches d'*E. coli*, sont réalisés. En effet, la conjugaison intraspécifique chez *E. coli* est relativement aisée à pratiquer et la croissance d'*E. coli* rapide.

Par exemple, lors du clonage des régions promotrices des gènes *bas* en amont du gène rapporteur *neo*, la sélection des exconjugants s'effectue via un simple crible positif qui vise à éliminer d'une part les cellules donneuses *E. coli* S17 dont le phénotype est Tet<sup>r</sup> et d'autre part les cellules receveuses *E. coli* SURE™ n'ayant pas acquis le plasmide recombinant qui confère la résistance à l'apramycine. Après sélection des exconjugants par les antibiotiques tétramycine et apramycine, un tapis cellulaire est observé au niveau des stries. Les exconjugants sont donc supposés avoir intégré les vecteurs recombinants. La capacité de transfert des plasmides recombinants est de ce fait démontrée. De plus, le fait que dans le témoin négatif (correspondant à un étalement d'*E. coli* SURE™ non transformée sur ce même milieu sélectif), aucune colonie ne se développe démontre l'absence de mutants spontanés résistants à l'apramycine

### **I.2.3 - CONJUGAISON INTERGENÉRIQUE *E. COLI*/ *S. AMBOFACIENS***

Lors des expériences d'analyse fonctionnelle des promoteurs *bas* par clonage en amont du gène codant l'EGFP, un témoin positif de conjugaison a été réalisé à l'aide de la souche *E. coli* S17 contenant le plasmide pIJ8630 recombinant ayant intégré le promoteur du gène *brdB* au site de clonage (MCS). Le gène *brdB* code le facteur sigma principal isolé chez *Streptomyces coelicolor*. Ce

gène est exprimé de façon constitutive. Un contrôle négatif constitué d'une souche *E. coli* S17 contenant le vecteur vide est également effectué. De nombreuses colonies de *Streptomyces* sont présentes sur toutes les boîtes après élimination des colonies d'*E. coli* par l'acide nalidixique et sélection des *Streptomyces* transformants pour la résistance à l'apramycine porté par le plasmide pIJ8630. Le phénotype des transconjugants de *Streptomyces ambofaciens* obtenus est de type sauvage c'est-à-dire que les colonies présentent un mycélium aérien (phénotype amy<sup>+</sup>) et une pigmentation grise (phénotype pig<sup>+</sup>). La différenciation de *S. ambofaciens* ne semble donc pas affectée par l'intégration du plasmide dans le chromosome bactérien.

Lors des expériences d'interruption des gènes *bas*, les souches sauvage et simples mutantes de *S. ambofaciens* sont transformées par conjugaison intergénérique avec *E. coli* ET12567/pUZ8002 contenant les cosmides 14C4 ou 25E1 modifiés, c'est-à-dire dans lesquels le gène *bas* a été remplacé par l'une des cassettes. Le plasmide pUZ8002 porte les fonctions de transfert nécessaires à la conjugaison mais n'est pas mobilisable puisque lui-même ne possède pas d'origine de transfert (Paget *et al.*, 1999).

Pour chaque construction, une culture d'*E. coli* ET12567/pUZ8002/cosmide modifié est amenée jusqu'à une DO<sub>600</sub> comprise entre 0,4 et 0,6 (la culture est réalisée en LB complété avec 25 µg/ml de chloramphénicol, 50 µg/ml de kanamycine et 25 µg/ml d'apramycine ou 25 µg/ml de spectinomycine, 37°C, 250 rpm). Tandis que ces cellules sont lavées deux fois avec du LB pour éliminer les antibiotiques qui inhiberaient la croissance de *Streptomyces* puis reprises dans 1/10 de volume initial de LB, environ 10<sup>8</sup> spores de *S. ambofaciens* pour chaque conjugaison sont diluées dans 0,5 ml de milieu de prégermination (2 × YT, (Kieser *et al.*, 2000)) et placées à 50°C pendant 10 min (choc thermique permettant la prégermination des spores). Un mélange entre 0,5 ml de cellules compétentes d'*E. coli* et 0,5 ml de spores est réalisé et des suspensions/dilutions de ce mélange sont étalées sur milieu SFM (MgCl<sub>2</sub> 10 mM) ou HT (MgCl<sub>2</sub> 10 mM) sans pression de sélection. Après incubation à 30°C pendant 16 à 20 h, les boîtes sont recouvertes par 1 ml d'eau stérile contenant 0,5 mg d'acide nalidixique (contre-sélection d'*E. coli*) et 1,25 mg d'apramycine ou de spectinomycine (sélection de l'intégration de la cassette sur le chromosome de *S. ambofaciens*) puis replacées à 30°C. Lorsque des colonies de *Streptomyces* sont visibles et sporulantes (après une semaine environ), des répliques sont effectuées sur HT acide nalidixique (25 µg/ml), apramycine ou spectinomycine (25 µg/ml) avec ou sans kanamycine (50 µg/ml). Les exconjugants apparus par double crossing-over (remplacement du gène *bas*) sont sensibles à la kanamycine [kan<sup>S</sup>] et résistants à l'apramycine [apra<sup>R</sup>] ou à la spectinomycine [spec<sup>R</sup>]. Les clones positifs (c'est-à-dire [apra<sup>R</sup> kan<sup>S</sup>] ou [spec<sup>R</sup> kan<sup>S</sup>]) sont sous-clonés deux fois de façon à obtenir une population homogène puis leur ADN est analysé par hybridation et PCR.

Pour chaque expérience de conjugaison, un témoin consistant en une conjugaison sans ajout de cellules d'*E. coli* ET12567/pUZ8002/cosmide modifié est réalisé. Ainsi, 0,5 ml de spores prégermées de *S. ambofaciens* sont étalés sur les différents milieux de conjugaison et permettent de dénombrer le taux de résistants spontanés aux antibiotiques considérés.

### I.2.4 - TRANSFORMATION DE PROTOPLASTES DE STREPTOMYCES

Lors des expériences d'analyse fonctionnelle des promoteurs *has* grâce au gène rapporteur néo, les vecteurs recombinants ont été transférés dans la souche hôte *S. lividans* ZX7 par protoplastisation. Des protoplastes de *S. lividans* ZX7 ont été préparés extemporanément à une concentration de  $3.10^9$  protoplastes par millilitre. Le taux de régénération est estimé par étalement de suspensions diluées de ces protoplastes sur milieu de régénération (R5) sans pression de sélection. Il est, dans notre expérience, de l'ordre de 1,5%. Par ailleurs, sur des protoplastes ( $1,5.10^8$  environ) régénérés sur milieu R5 pendant 20 h, est coulée une solution de thiostrepton (500  $\mu\text{g/ml}$ ). Après séchage de la boîte, une concentration finale de 50  $\mu\text{g/ml}$  est obtenue. Cette expérience permet de calculer la fréquence de résistants au thiostrepton apparus spontanément. Cette fréquence est inférieure à  $10^{-7}$ . Enfin, la transformation avec le plasmide natif (50 ng) permet d'estimer l'efficacité de transformation qui s'élève à  $5.10^4$  transformants par microgramme d'ADN. Les transformations avec les mélanges de ligation (5  $\mu\text{l}$  du volume final de ligation) ont produit environ 1200 transformants pour le mélange de ligation réalisé avec l'ADN de p1338 et 203 avec celui impliquant l'ADN de p1329. L'ensemble de ces résultats est regroupé dans le Tableau MM-3.

Mélange de ligation	Transformation de <i>S. lividans</i> ZX7	
	p1J487/p1338	p1J487/p1329
Pourcentage de régénération	1,5%	1,5%
Fréquence de résistants spontanés au thiostrepton	$< 10^{-7}$	$< 10^{-7}$
Efficacité de transformation (transformants/ $\mu\text{g}$ d'ADN)	$5.10^4$	$5.10^4$
Nombre de transformants obtenus	1200	203
Nombre de transformants analysés	22	19
Nombre de transformants renfermant un vecteur recombinant	0	2

Tableau MM-3 : Expériences de transformation de protoplastes de *S. lividans* ZX7.

## **I.3 ☞ OBSERVATION DES COLONIES BACTERIENNES A LA RECHERCHE D'UN PHENOTYPE**

### **I.3.1 - OBSERVATIONS MORPHOLOGIQUES**

Les observations morphologiques des souches sauvage et doubles mutantes sont réalisées sur différents milieux dont la composition est résumée dans le Tableau MM-2. Les photographies sont réalisées par un appareil photonumérique (Nikon D1) couplé à une loupe binoculaire.

### **I.3.2 - TRAITEMENTS AUX AGENTS OXYDANTS CHIMIQUES**

Les stress oxydatifs ont été réalisés selon la méthode de Paget *et al.* (1999). Une quantité de  $10^6$  ou  $10^7$  spores est ajoutée à 4 ml de milieu SNA en surfusion. Après homogénéisation, ce mélange est déposé sur des boîtes de Petri contenant précisément 25 ml de milieu NA préalablement séché 30 minutes. Après prise en masse du milieu SNA et séchage des boîtes pendant 5 min, des disques de papier Whatman (5 mm de diamètre) sont déposés sur le tapis bactérien puis imbibés avec 10  $\mu$ l d'agent stressant ( $H_2O_2$ , diamide ou ménadione) à différentes concentrations. Les boîtes sont incubées pendant 24 h à 30°C. Un gradient de concentration est généré par diffusion de l'agent oxydant dans le milieu à partir du disque. Suite à ce traitement, il se délimite autour du disque, sur le tapis confluent, une zone d'inhibition de croissance dont le diamètre est mesuré.

### **I.3.3 - TRAITEMENT AUX RAYONNEMENTS ULTRAVIOLETS (UV)**

Des suspensions/dilutions de spores du sauvage et des simples et doubles mutants sont étalées sur milieu HT puis irradiées dans les mêmes conditions par des radiations UV émises par deux lampes (GTE Sylvania). Les doses d'UV (0, 100, 200, 300 et 400  $J/m^2$ ) sont mesurées précisément avec un radiomètre (ITT Cannon, sonde à 254 nm). Les boîtes sont incubées à l'obscurité afin de limiter le processus de photoréactivation (Volf *et al.*, 1993). Le taux de survie pour chaque dose d'UV est calculé en rapportant le nombre d'UFC obtenu après traitement sur celui obtenu sans traitement.

Pour ces deux traitements (agents oxydants chimiques et UV), les échantillons sont effectués en duplicats, voire triplicats, afin d'établir un intervalle de confiance à chaque valeur. L'intervalle de

confiance est calculé selon la formule du logiciel Excel (Microsoft) avec une probabilité de 95% de trouver la moyenne des échantillons dans cet intervalle. Les courbes sont effectuées à l'aide du logiciel Excel.

#### **I.3.4 - DETERMINATION DE L'ACTIVITE ANTIBACTERIENNE**

L'activité antibactérienne des souches mutantes des gènes *bas* a été testée en milieu solide HT et R2. Sur boîtes de Petri contenant 25 ml de milieu solide HT ou R2,  $10^7$  spores des souches sauvage et doubles mutantes sont étalées en tapis confluent. Les boîtes sont incubées à 30°C et aux jours 2, 4, 6 et 8, un cylindre de milieu est prélevé et conservé à 4°C. Puis, ces cylindres sont déposés sur boîtes HT où a été précédemment étalé, en tapis confluent, la souche indicatrice *Bacillus subtilis*. Afin de permettre la diffusion dans le milieu des composés à activité antibactérienne à partir des cylindres, les boîtes sont incubées 6 h à 4°C puis 12 h à 37°C. Une zone d'inhibition autour du cylindre indique une activité antibactérienne.

- CHAPITRE II -

MANIPULATION DES ACIDES  
NUCLEIQUES



## II.1 ☞ PREPARATION DES ACIDES NUCLEIQUES

### II.1.1 - PREPARATION D'ADN, DIGESTION ENZYMATIQUE ET ELECTROPHORESE

#### II.1.1.1 - EXTRACTION ET PURIFICATION D'ADN

L'ADN total de *Streptomyces* est extrait selon la méthode mise au point au laboratoire (Leblond *et al.*, 1996). L'ADN plasmidique de *Streptomyces* est extrait par une méthode dérivée de la lyse alcaline (Ish-Horowicz and Burke, 1981) adaptée à la forme de croissance mycélienne de la bactérie (Hopwood *et al.*, 1985). La préparation d'ADN de *Streptomyces* pour les électrophorèses en champs pulsés ou PFGE est réalisée suivant la méthode décrite par Leblond *et al.* (1996) et les migrations sont faites dans une cuve Bio-Rad de type CHEF ("Contour clamped Homogeneous Electric Field") (Chu *et al.*, 1986).

Pour les échantillons d'ADN destinés à l'amplification en chaîne (PCR), l'ADN est au préalable purifié par extraction phénol/chloroforme (v/v) (Sambrook *et al.*, 1989). Pour la préparation du vecteur de clonage pIJ487, l'ADN plasmidique est purifié sur gradient de chlorure de césium (CsCl, 1mg/ml) en présence de bromure d'éthidium (BET, concentration finale 0,9 mg/ml) (Sambrook *et al.*, 1989). L'ADN des bactériophages M13 recombinants, en vue des sous clonages, est extrait sous sa forme double brin à partir de cultures d'*E. coli* JM101 par la méthode de lyse alcaline (Ish-Horowicz and Burke, 1981). Le dosage de l'ADN sur gel est réalisé selon Sambrook *et al.* (1989).

L'extraction de l'ADN simple brin de bactériophages M13 recombinants est réalisée après 4 h de culture à 37°C de la souche d'*E. coli* infectée par un bactériophage M13 recombiné. Le surnageant de la culture est additionné de PEG 6000 (Polyéthylène glycol) à la concentration finale de 2% et de NaCl (concentration finale 0,25 M). Après centrifugation (12.000 rpm, 5 min), le culot est redissout dans 100 µl de tampon TE (Tris-HCl 10 mM, Ethylenediaminetetraacetic acid ou EDTA 1 mM, pH 8,0) puis traité par une extraction phénol (pH 8,0) / chloroforme (v/v). L'ADN est ensuite précipité, lavé à l'éthanol 70%, séché et repris dans du tampon TE. L'ADN est alors de qualité requise pour le séquençage automatique.

#### II.1.1.2 - DIGESTIONS PAR LES ENDONUCLEASES ET ELECTROPHORESE D'ADN

Les enzymes de restriction sont utilisées selon les recommandations du fournisseur (Roche Molecular Biochemicals). Les électrophorèses en gel d'agarose sont réalisées en tampon TAE

(Tris-acétate 0,04 M ; EDTA 0,001 M) ou TBE (Tris-Borate 0,09 M ; EDTA 0,002 M). Le voltage et la durée de migration sont adaptés pour une résolution optimale des fragments d'ADN. Le système de traitement d'images GelDoc 1000 (Biorad) permet de visualiser les gels d'électrophorèse après immersion dans une solution de BET à 0,5 µg/ml et de stocker le résultat. L'ADN du bactériophage λ digéré par *Pst*I est utilisé comme standard de migration (Daniels *et al.*, 1983).

La séparation des fragments d'ADN d'amplification *has* (couple d'amorces L1/B et R1/B) ne différant que par quelques nucléotides (558 pb et 567 pb), une électrophorèse en gel d'agarose à 4% (1/3 :2/3 LMPA Eurobio, SeaKem® FMC) en tampon TBE 0,5 x à 10 V/cm a été réalisée (Sambrook *et al.*, 1989).

### II.1.1.3 - LIGATION

Lors des différents clonages réalisés au cours de ce travail, les ADN des vecteurs de clonage sont préalablement déphosphorylés et les mélanges de ligation sont réalisés avec un ratio ADN insert/ADN vecteur de 10 (par exemple, 30 ng et 3 ng de vecteur dans un volume final de 10 µl). Un témoin, correspondant à un aliquot (1 µl) du mélange de ligation avant addition de la ligase est déposé sur gel, en parallèle d'un aliquot du mélange de ligation après action de la ligase (1 µl). L'efficacité de la ligation est déduite de l'observation de bandes de poids moléculaire accru après ligation par comparaison au témoin.

La Figure MM-10 présente l'exemple du contrôle électrophorétique de l'efficacité de ligation entre le vecteur pIJ487 et les bactériophages recombinants pL329 et pL338, lors du clonage des régions promotrices des gènes *hasL* et *hasR* respectivement en amont du gène rapporteur néo (voir paragraphe des Résultats B-III.3.1).

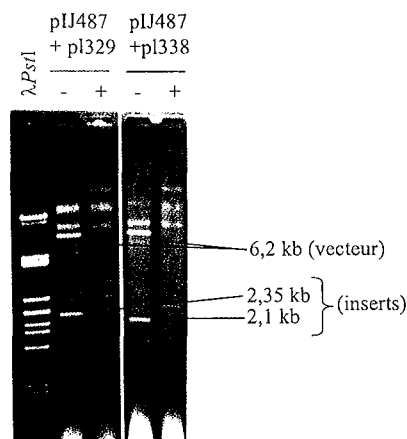


Figure MM-10 : Contrôle électrophorétique des mélanges de ligation entre pIJ487 et les bactériophages recombinants pL329 et pL338.

Les pistes (-) illustrent les mélanges de ligation de pIJ487/pL329 et pIJ487/pL338 (digéré par *Sac*I) avant ajout de ligase. Les pistes (+) illustrent les mêmes mélanges après ligation. Les conditions de ligation et de migration sont indiquées dans le matériel et méthodes.

### II.1.2 - PRÉPARATION ET DOSAGE D'ARN

Les ARN totaux sont extraits à partir de mycélium cultivé en milieu HT (culture de 10 ml). Le mycélium est resuspendu dans 2 ml de Tri Reagent™ (Sigma) et lysé par sonication (80 W pendant deux périodes de 5 s) sur la glace. Le Tri Reagent™ est un mélange de deux composants principaux ; le thiocyanate de guanidine et le phénol. Après addition de chloroforme au lysat cellulaire, les ARN sont en solution dans la phase aqueuse, l'ADN dans l'interphase, et les protéines dans la phase organique. La phase aqueuse est alors réservée, puis une précipitation à l'isopropanol est réalisée permettant de concentrer les ARN par centrifugation 5 min à 12.000 rpm. Le culot est alors lavé à l'éthanol 70% (v/v, éthanol absolu et eau traitée au DEPC diéthyl pyrocarbonate, (Sambrook *et al.*, 1989)) puis repris dans un volume d'eau DEPC.

La concentration d'ARN est estimée par spectrophotométrie par le rapport de densité optique 260 nm/280 nm après dilution au 100<sup>ème</sup> de la solution. Ce rapport doit être compris entre 1,8 et 2 pour garantir une pureté compatible avec l'action des enzymes utilisés par la suite. La concentration d'ARN ( $\mu\text{g/ml}$ ) est donnée en multipliant la valeur de la DO à 260 nm par 40 et en tenant compte du facteur de dilution (Sambrook *et al.*, 1989). La qualité et la concentration peuvent également être estimées par électrophorèse en gel d'agarose contenant du formaldéhyde (concentration finale 2,2 M). Le tampon de migration consiste en une solution de tampon MOPS 20 mM, EDTA 1 mM et acétate de sodium 5 mM (Sambrook *et al.*, 1989). Le marqueur de taille utilisé est le génome à ARN du virus Brome Mosaic (BMV RNA, Promega). Le gel après électrophorèse est immergé dans une solution de Radiant™ Red (dilution au 1000<sup>ème</sup> de la solution commerciale Biorad).

## II.2 Les techniques utilisant les acides nucléiques

### II.2.1 – HYBRIDATION ADN - ADN

L'ADN utilisé comme sonde est marqué à l'aide du kit "DIG DNA Labeling et Detection Kit Nonradioactive" (Roche Molecular Biochemicals) par incorporation du DIG-dUTP (déoxyUridine TriPhosphate couplé à la digoxigénine) par la méthode des amorces aléatoires.

Après migration en gel d'agarose, l'ADN est transféré sur membrane de Nylon de type N (Schleicher et Schuell) à l'aide du système Vacugene (Pharmacia-LKB). Trois traitements successifs sont appliqués : dépurination (solution HCl 0,25 N) pendant 20 min, dénaturation (solution NaOH 0,5N ; NaCl 1,5 M) pendant 20 min et transfert (solution SSC 20X : NaCl 3 M, citrate de Na 0,3 M) pendant 1 h. Rincée dans du SSC 2X, la membrane est ensuite séchée. L'ADN est fixé à la membrane par exposition à un rayonnement UV ( $\lambda=254$  nm) pendant 5 min. Les étapes de pré-hybridation, hybridation, lavages des membranes ainsi que la révélation ont été réalisées d'après les instructions du kit "DIG DNA Labeling et Detection Kit Nonradioactive" (Roche Molecular Biochemicals). L'hybridation ainsi que les lavages stringents sont réalisés à 68°C. Cette température permet la détection des duplex d'ADN présentant une identité de séquence proche de 100%. La détection de signaux est réalisée grâce à une caméra refroidie par effet Peltier (Fluor's max, Biorad) couplée à une carte d'acquisition de données gérée par le logiciel Quantity One™ (Biorad).

### II.2.2 – SEQUENÇAGE AUTOMATIQUE D'ADN

Le séquençage de l'ADN est réalisé selon la méthode de Sanger *et al.* (1977). La détection du signal est réalisée grâce à l'incorporation de didéoxynucléotides couplés à une molécule fluorescente. Le kit "Dye terminator cycle sequencing" (Perkin-Elmer Applied Biosystem) est utilisé. Les produits d'élongation sont séparés par électrophorèse capillaire en gel de poly N,N-diméthylacrylamide et analysés automatiquement par le système ABI Prism 310. Les données sont collectées et analysées par le logiciel Sequencing analysis (Perkin-Elmer Applied Biosystem).

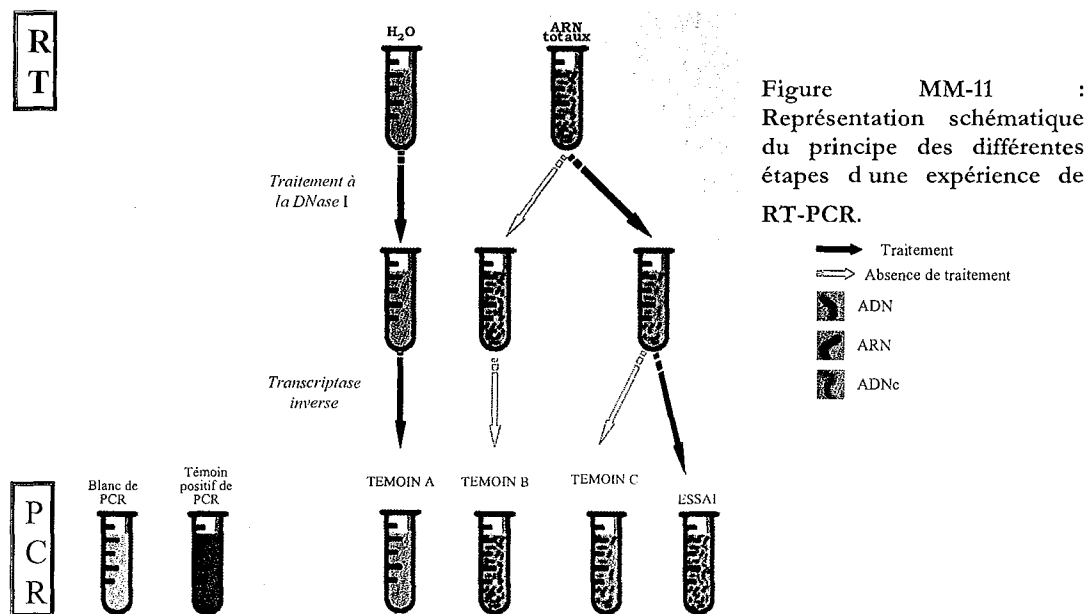
### II.2.3 – TRANSCRIPTION INVERSE (RT) ET AMPLIFICATION D'ADN (PCR)

#### II.2.3.1 – DIFFICULTES DE LA METHODE DANS LE CAS DES BACTERIES

Les approches de RT-PCR développées chez les eucaryotes supérieurs tirent parti de la présence d'introns pour distinguer l'ADN complémentaire (ADNc, produits de la transcriptase inverse)

d'une contamination d'ADN génomique. En effet, l'élimination des introns par épissage des ARN pré-messagers rapproche les séquences exoniques. Ainsi, le choix judicieux de la localisation des amorces de PCR dans les exons permet d'obtenir un produit d'amplification à partir des ADNc et non de l'ADN génomique contaminant. En revanche, l'absence d'intron chez les bactéries rend indispensable l'élimination de toute trace d'ADN chromosomique contaminant entraînant un faux positif dans la détection des transcrits en RT-PCR.

La transcription inverse des ARN totaux comprend trois étapes majeures : la fixation d'amorces aléatoires (hexanucléotides) sur les ARN, le traitement à la DNase I et la transcription inverse (Figure MM-11).



Trois témoins sont réalisés : (i) un témoin négatif de transcription inverse appelé blanc de RT. Il est obtenu en traitant de l'eau, par ailleurs utilisée dans les mélanges réactionnels de l'expérience, par la DNase I et la transcriptase inverse (témoin A) ; (ii) un échantillon d'ARN traités ni par la DNase I ni par la transcriptase inverse. Ce témoin permet de mettre en évidence la présence d'ADN contaminant dans les échantillons d'ARN (témoin B) ; (iii) un témoin correspondant à un aliquot de l'essai prélevé juste après le traitement à la DNase I et non traité par la transcriptase inverse. Il permet de vérifier l'efficacité de l'élimination de l'ADN contaminant (témoin C). L'essai est constitué d'un échantillon d'ARN traité à la DNase I et par la transcriptase inverse.

L'amplification spécifique, par PCR, des ADNc générés par l'étape de transcription inverse est obtenue grâce à l'utilisation d'amorces (19-mers) spécifiques des séquences des ORF étudiées (Tableau MM-6). L'essai ainsi que les témoins réalisés lors de l'étape RT sont soumis à une amplification par PCR. Deux témoins de PCR sont réalisés : un témoin négatif, appelé blanc de PCR, correspondant à une réaction de PCR sur un volume d'eau (certifiant l'absence de contamination des réactifs) et un témoin positif réalisé sur un échantillon d'ADN génomique de *S. ambifaciens*, souche sauvage, de quantité déterminée (10 pg). La Figure MM-12 présente un exemple de résultat mettant en évidence la transcription de *recA*.

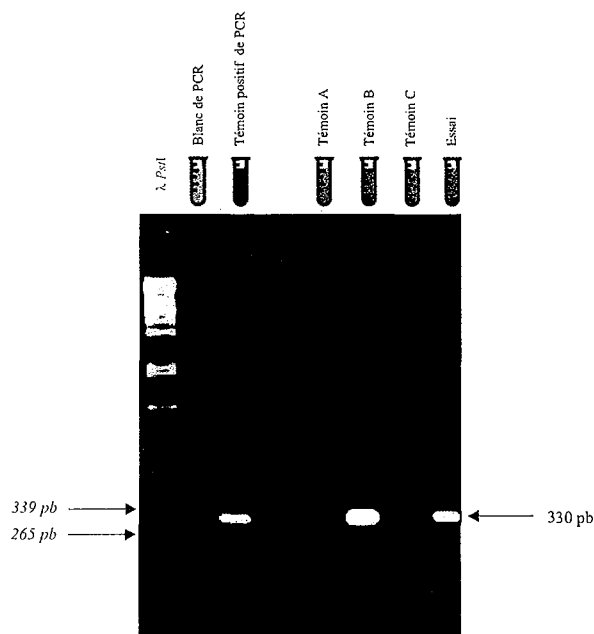


Figure MM-12 : Electrophorèse des produits d'amplification spécifique des ADNc des transcrits *recA* chez la souche sauvage *S. ambfaciens* DSM40697.

Afin d'être présentées en cohérence avec la Figure MM-11, les pistes d'électrophorèse, provenant de la même expérience, ont été découpées puis réordonnées. Les conditions d'électrophorèse sont indiquées dans le matériel et méthodes. Le symbolisme et le principe des témoins en haut de la Figure sont identiques à la Figure MM-11. Les ARN ont été extraits après 16 h de croissance en milieu liquide.

### II.2.3.2 – CHOIX D'UN TÉMOIN INTERNE POUR UNE APPROCHE QUALITATIVE : LE GÈNE *RECA*

Le gène *recA* a été choisi comme témoin interne de l'expérience de RT-PCR. En effet, ce gène codant la protéine majeure de la recombinaison homologue et de l'induction du système SOS était supposé être exprimé dans toutes les phases de croissance. Il constitue donc un bon candidat comme témoin positif d'expression dans une approche qualitative de détection de la transcription par RT-PCR. La séquence du gène *recA* de *S. ambfaciens* DSM40697 avait été déterminée au laboratoire permettant le choix d'un couple d'amorces A1/A2 localisées à l'intérieur de la phase codante (Tableau MM-6, (Aigle *et al.*, 1996)).

L'analyse des résultats en terme semi-quantitatif n'est pas envisagée compte tenu de la référence disponible. Ainsi, il faut disposer d'un témoin interne dont le niveau d'expression est connu à tous les points de l'observation. En effet, pour comparer les intensités des signaux obtenus, il faut pouvoir rapporter l'intensité des signaux de PCR obtenus à une référence. C'est à cette seule condition que les fluctuations de l'intensité relative des signaux peuvent être interprétées en terme de modulation de l'expression génique. Cette condition n'est pas remplie avec le gène *recA*. En effet, le niveau d'expression de *recA* au cours des phases de la croissance n'est pas connu. De

plus, elle est induite lors de stress, notamment en réponse à des agressions du matériel génétique (système SOS). Enfin, chez les *Streptomyces*, l'expression de *recA* est fortement modulée par le contexte génétique, particulièrement chez les mutants issus de l'instabilité (Aigle *et al.*, 1996).

### II.2.3.3 – LES MILIEUX REACTIONNELS DE RT-PCR

#### II.2.3.3.1 – La transcription inverse des ARN en ADNc

Les différentes étapes nécessaires à la synthèse d'ADNc à partir d'ARN totaux isolés précédemment sont présentées sur la Figure MM-13 et la composition des mélanges réactionnels dans le Tableau MM-4.

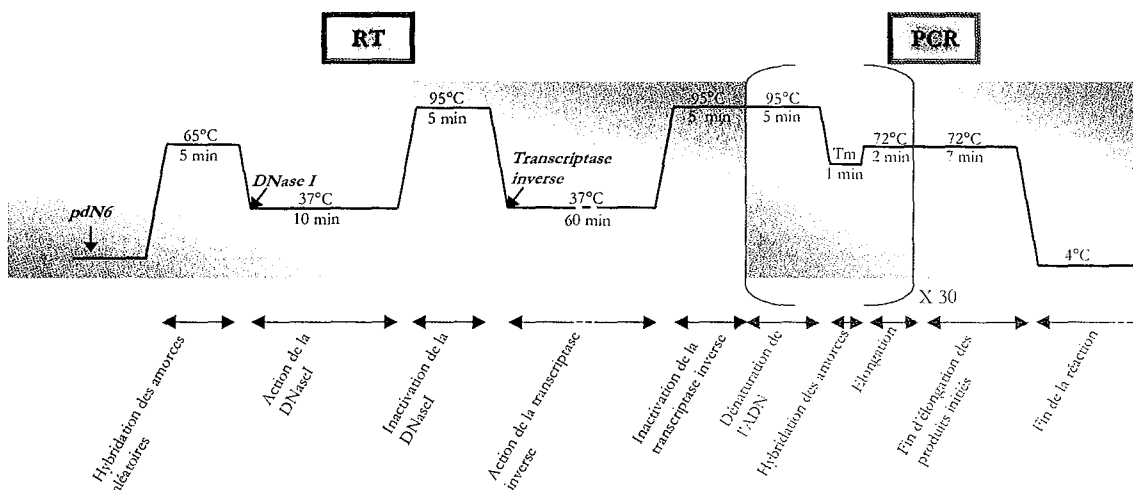


Figure MM-13 : Représentation des cycles de température utilisés dans les expériences de RT-PCR.

Les flèches rouges indiquent l'ajout d'un composant au mélange réactionnel. Les flèches horizontales indiquent, schématiquement, les étapes des différentes réactions.  $T_m$  correspond à la température spécifique d'hybridation des amorces de PCR. Les hexanucléotides aléatoires utilisés en transcription inverse (RT) sont indiqués par pdN6.

Les ARN totaux sont extraits à partir de 2 cultures indépendantes pour chaque souche étudiée et les ADNc sont fabriqués à partir de 2 lots d'ARN extraits d'une même culture, et pour les 3 points de la croissance. Un échantillon d'ARN (2 à 4  $\mu$ g) est incubé avec des amorces aléatoires (hexanucléotides de synthèse de séquences aléatoires, pdN6, Pharmacia) dans un volume final de 5  $\mu$ l pendant 5 min à 65°C. Les pdN6 vont alors s'ancrer sur les acides nucléiques et permettront l'amorçage de la transcriptase inverse. Le mélange réactionnel (20  $\mu$ l final) est composé : de tampon de transcription inverse concentré 1 x (tampon 5x : Tris 50 mM pH 8,3, DTT ou dithiothreitol 10 mM,  $MgCl_2$  8 mM), de dNTP (Roche Molecular Biochemicals), d'un inhibiteur de RNase d'origine porcine (RNA Guard®, Pharmacia) et de DNase I (Pharmacia). L'incubation à 37°C 10 min. permet ainsi la dégradation de l'ADN génomique contaminant les ARN. Le

passage à 95°C dénature la DNase I. Le mélange réactionnel est alors additionné de transcriptase inverse (Moloney Murine Leukemia Virus Reverse Transcriptase, enzyme appelé ultérieurement RT, Pharmacia). Le mélange est incubé à 37°C pendant 60 min.

Réactif	Concentration finale
Hexanucléotides aléatoires (pdN6)	100 ng/ $\mu$ l
Tampon de RT	1 X
dNTP	20 $\mu$ M
RNA Guard $\text{\textcircled{R}}$	1 UE/ $\mu$ l
DNase I	0,25 UE/ $\mu$ l
M-Mu-LV Reverse Transcriptase	0,16 UE/ $\mu$ l

Tableau MM-4 : Composition du mélange réactionnel utilisé dans les expériences de transcription inverse (RT) des ARN totaux de *S. ambofaciens*.

### II.2.3.3.2 – L'amplification d'ADN par polymérisation en chaîne (PCR)

La PCR ou réaction de polymérisation d'ADN en chaîne consiste en l'amplification de manière exponentielle d'un segment d'ADN à l'aide de deux amorces complémentaires des séquences flanquant l'ADN à amplifier (Mullis *et al.*, 1986). La polymérase d'ADN utilisée est la Taq polymérase (*Thermus aquaticus*) dans les conditions spécifiées par le fournisseur (Appligène). Le mélange réactionnel est indiqué dans le Tableau MM-5.

Réactif	Concentration finale
Amorces	25 pmoles de chaque
Tampon de PCR	1 X
dNTP	1 mM
Taq polymérase	0,04 UE/ $\mu$ l
DMSO	2,5 à 5%

Tableau MM-5 : Composition du mélange réactionnel utilisé dans les expériences de polymérisation en chaîne (PCR) chez *S. ambofaciens*.

Le tampon de PCR est composé de Tris HCl 10 mM pH 9, KCl 50 mM, MgCl<sub>2</sub> 1,5 mM, Triton X100 0,1% et Bovine Serum Albumine (BSA) 0,2 mg/ml. Le témoin négatif consiste en une réaction sans matrice réalisée sur un volume d'eau ultra-pure. Le cycle de température utilisé est représenté sur la Figure MM-13.

Les amorces utilisées, oligonucléotides de synthèse (MWG-Biotech GmbH) sont choisies pour assurer une composition en bases G+C équivalente et garantir des températures de fusion ou T<sub>m</sub> équivalents (l'ensemble des amorces utilisées dans les réactions de PCR au cours de ce travail est résumé dans le Tableau MM-6).



Utilisation	Amorce	Séquence
<u>RT-PCR</u>	A	AGGTGGCGGACCCGTCCAA
	B	ATGAGCCTGCCGGAGGACG
	D	GGCGCCAACGGCTACAAC
	R1	GGCCGATGAGGCCGGCCGT
	R2	GACCAC TGAGCGCCGCTCA
	L1	CGCTGCTGACGAGCAGGCA
	L2	GAAAGGCAGTGGCCGGCGG
	A1	TCCTGTCCCAGCCGGACAA
	A2	GTGCGACCGAGGCGTAGAA
	A1	TCCTGTCCCAGCCGGACAA
	A2	GTGCGACCGAGGCGTAGAA
	<u>RT-PCR quantitative</u>	Rqu1
Rqu2		GATGTTCCCTCCCAGTCGATCGG
Rqu3		CCGACGAGACTGCGCTTAC
Rqu4		GGCGAGACAGACCGAGAAAA
Lqu1		CTGCTGACGAGCAGGAG
Lqu2		GATGGGCCCTCCCTGTT
Lqu3		CCCGAGGCCGGGGTATTG
HrdB- HrdB-		CGCGAGCCCATCTCGCTGC CGTCGAGGCTCTTCGGCTG
<u>Extension d'amorces et sondes des expériences de cartographie à la nucléase S1</u>	R5	GTCATCCACGGTGGTGGCAGCGGCA
	G	GTCTTGGGCGCCACCTGGACGGG
	E	GGCATGCCCGTGGTGTGTCTCGG
	L5	CAGCACAGCGTGTACGTGGTGTCC
	arsR1	CGGTGTCAACGTAGACGGA
	arsR2	TCTCCACTGGCCCGGATAC
<u>Construction des mutants <math>\Delta</math>has</u>	HasL1	GTGACGAGCAGGCAGTTCGAACAGGGAGGGCCCATCATGTGTAGGCTGCAGCTGCTTC
	HasL2	GCCGCCCCGGTGTGCCACGGTGGGGGCCCTCGCCTCAATTCCGGGGATCCGTCGACC
	HasR1	GCCGTGAGCGGGGCCGATCGACTGGGAGGGAACATCATGTGTAGCCTGGACCTCCTTC
	HasR2	ACCGCCGCGGAAGGACAGCCGCCCTGAGCGGCGTCAATTCCGGGGATCCGTCGAC
	CL1	GTCTACACCCCCCCCCG
	CL2	GCCACGTACCCCCCGGT
	CR1	ACACGACGGACGACGGCA
	CR2	GGAGCGACCTGCGGGAGA
<u>Construction des complémentants</u>	CoR1	TATGAATCTGTGACGGTGTGCGGATG
	CoR2	AATGAATFCCCGGGCGTGCAGGATGA
	CoL1	TATGAATFCGACCGGTGCGTGTGAGGC
	CoL2	ATTGAATFCGCGGACGTCTCCCGGCC

Tableau MM-6 : Amorces utilisées dans les réactions de PCR au cours de ce travail.

L'utilisation du logiciel DNA Amplify 2.5 (University du Wisconsin) permet de valider le choix des amorces : spécificité d'amplification du segment d'ADN choisi (dans la séquence soumise), faible probabilité de dimérisation des amorces, absence de formation de repliment au sein d'une amorce.

Pour les expériences de RT-PCR, les matrices correspondent à 2µl d'ADNc produits lors de l'étape de transcription inverse.

Dans le cas de l'amplification des cassettes Apra et Spec lors de la construction des mutants interrompus pour le(s) gène(s) *has*, les matrices correspondent à l'ADN purifié des fragments *EcoRI-HindIII* des plasmides pIJ773 et pIJ778 respectivement portant les cassettes de résistance. Deux températures d'appariement des amorces sont utilisées dans les réactions d'amplification des cassettes (Figure MM-14) : en effet, une première série de 10 cycles à une température de 50°C permet l'appariement de la partie des amorces spécifique de la cassette (séquences à bas taux en bases G et C) puis une seconde série de 15 cycles à la température de 55°C permet à la totalité des amorces, et notamment la partie spécifique du génome de *Streptomyces* (fort taux en bases G et C), de s'hybrider. Ainsi, une amplification efficace de la totalité de la cassette flanquée des séquences spécifiques des loci *has* a été obtenue.

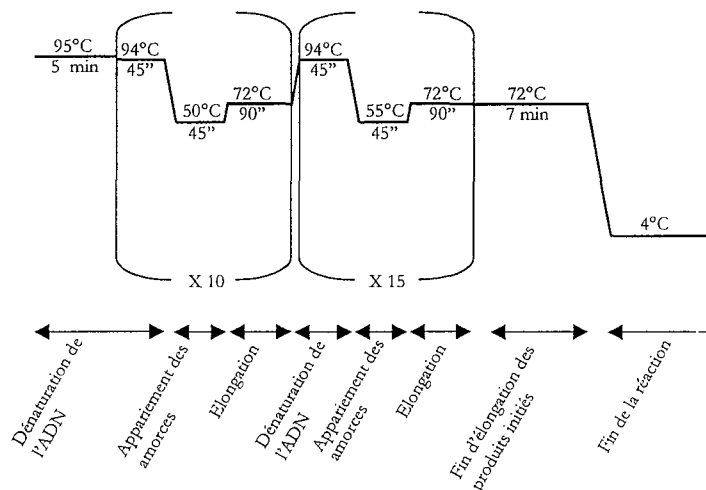


Figure MM-14 : Représentation des cycles de températures utilisés dans les expériences d'amplification des cassettes Apra et Spec lors de la construction des mutants interrompus pour le(s) gène(s) *has*.

Les flèches horizontales indiquent schématiquement les étapes des différentes réactions.

Les expériences de PCR quantitative en temps réel sont détaillées dans le chapitre III.

### II.2.4 - EXTENSION D'AMORCES

Le principe de l'identification des sites d'initiation de la transcription par des expériences d'extension d'amorces est résumé sur la Figure MM-15. Les ARN totaux sont extraits à partir d'au moins 2 cultures indépendantes pour chaque souche étudiée et pour chaque condition (phase de croissance en milieu liquide ou en milieu solide, conditions de stress...). Des échantillons de 40µg d'ARN sont ensuite incubés avec une amorce spécifique (voir Tableau MM-6) préalablement marquée au <sup>32</sup>P (selon Kieser *et al.*, 2000), puis traités à la transcriptase inverse selon Kieser *et al.* (2000). Les produits de transcription inverse ainsi obtenus sont déposés sur un gel de polyacrylamide 8%, en parallèle des produits de réactions de séquençage réalisées avec la même amorce. Après migration, séchage et transfert, le gel est exposé à un film autoradiographique à -70°C pendant 12 à 150h environ selon les cas.

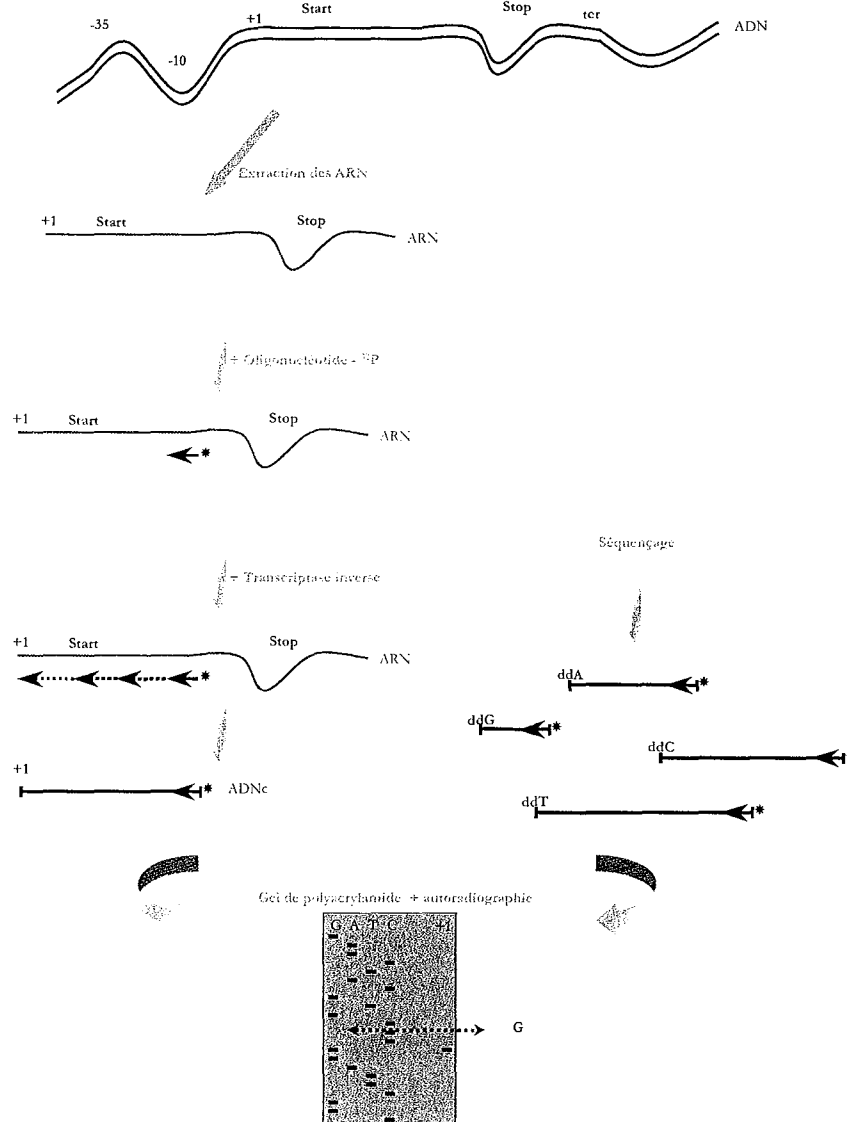


Figure MM-15 : Principe de la détection de sites d'initiation de la transcription par extension d'amorces.

### II.2.5 - CARTOGRAPHIE A LA NUCLEASE S1

Le principe de l'identification des sites d'initiation de la transcription par des expériences de cartographie à la nucléase S1 est résumé sur la Figure MM-16. Comme pour les expériences d'extensions d'amorce, les ARN totaux sont extraits à partir d'au moins 2 cultures indépendantes pour chaque souche étudiée et pour chaque condition. Les sondes utilisées dans ces expériences sont réalisées par PCR (voir Tableau MM-6), l'oligonucléotide anti-sens ayant préalablement été marqué au  $^{32}\text{P}$ . Des échantillons de  $40\mu\text{g}$  d'ARN sont ensuite incubés avec cette sonde puis traités à la nucléase S1 selon Kieser *et al.* (2000). Les produits de digestion par la nucléase S1 ainsi obtenus sont déposés sur un gel de polyacrylamide 8%, en parallèle des produits de réactions de séquençage réalisées avec la même amorce anti-sens, que celle utilisée pour fabriquer la sonde. Après migration, séchage et transfert, le gel est exposé à un film autoradiographique à  $-70^\circ\text{C}$  pendant 12 à 150h environ selon les cas.

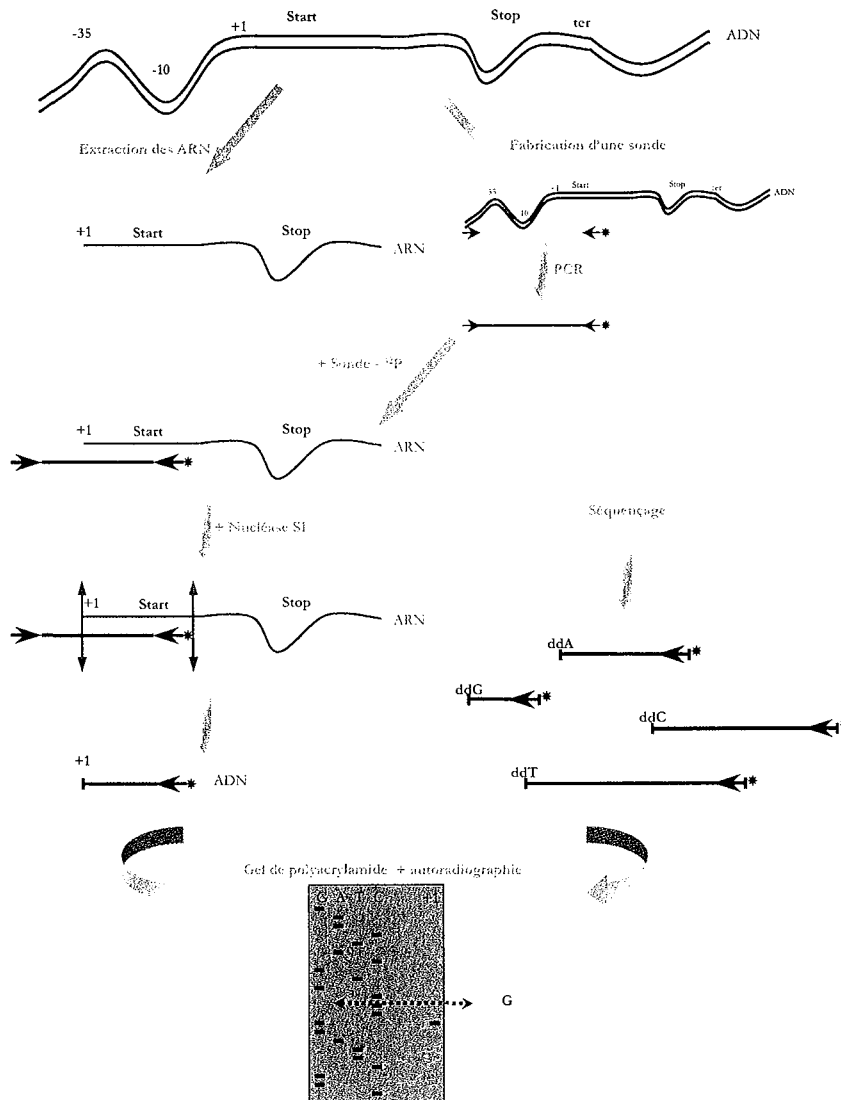


Figure MM-16 : Principe de la détection de sites d'initiation de la transcription par cartographie à la nucléase S1.

## II.3 $\propto$ ANALYSE DES SEQUENCES D'ADN

### II.3.1 - DETECTION DE PHASES OUVERTES DE LECTURES

L'identification des cadres de lecture ouverts ou ORF (Open Reading Frame) est réalisée par le logiciel Framplot 2.3 (Bibb *et al.*, 1984). La présence de structures secondaires de type tige-boucle a été recherchée à l'aide du logiciel DNA mfold 3.0 (SantaLucia, 1998) (<http://mfold2.wustl.edu>). Les séquences consensus ont été localisées à l'aide du logiciel DNA Strider 1.2 (Marck, 1988). Les programmes "BLAST" (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990) ont été utilisés pour rechercher les homologues des séquences déterminées avec les séquences nucléotidiques et protéiques des banques de données (via le National Center for Biotechnology Information, NCBI, <http://www.ncbi.nlm.nih.gov/>) et le Centre Sanger (<http://www.sanger.ac.uk>) qui réalise le séquençage du génome de *S. coelicolor* A3(2) (sous la responsabilité de Sir Pr. D.A. Hopwood). Les domaines protéiques ont été recherchés dans les banques CDD, pfam, et SMART. Les alignements de séquences ont été réalisés par le programme SeqApp, sous programme Clustal V (Higgins and Sharp, 1988).

### II.3.2 - ANALYSE DE LA REDONDANCE D'UN GENOME PAR LE LOGICIEL BLASTCLUST

Le programme "BLAST-CLUST" ("Basic Local Alignment Search Tool - score-based single-linkage clustering") (Altschul *et al.*, 1990; Altschul *et al.*, 1997) a été utilisé pour déterminer le taux de redondance de certains génomes séquencés (programme accessible sur le site du "National Center for Biotechnology Information" (NCBI, <http://www.ncbi.nlm.nih.gov/>) et séquences protéiques disponibles sur les sites de "Institute for Genomic Research" (TIGR, <http://www.tigr.org/>) ou de "Sanger Institute" (<http://www.sanger.ac.uk/>).

Ce programme regroupe les séquences protéiques en familles (groupes de protéines comportant au moins 2 membres) selon les identités protéiques estimées par l'algorithme du logiciel BLAST. Deux paramètres doivent être définis :

- le seuil de similarité (S) : dans toutes les analyses réalisées au cours de ce travail ce seuil a été fixé à 50% ce qui correspond à une similarité minimale de 50% entre les protéines regroupées au sein d'une même famille.

- le seuil correspondant à la longueur minimale de recouvrement : dans toutes les analyses réalisées ici ce seuil a été fixé à 0,9, ce qui correspond à une longueur minimale de chevauchement de 90% entre les différentes protéines regroupées.

**- CHAPITRE III -**

**LA PCR QUANTITATIVE**

## III.1 $\infty$ PRINCIPE DE LA PCR QUANTITATIVE

### III.1.1 – LES BASES DE LA PCR QUANTITATIVE EN TEMPS REEL

Depuis les premiers travaux publiés en 1986, la PCR est devenue un outil essentiel de biologie moléculaire (S., 1999). A partir d'une simple copie d'une séquence particulière d'acides nucléiques, la PCR permet d'amplifier et donc de détecter spécifiquement cette séquence.

La PCR quantitative en temps réel est basée sur la nature exponentielle de cette technique.

En effet, la réaction de PCR peut être divisée en trois phases :

- *la phase de latence*,
- *la phase exponentielle* de synthèse qui démarre quand suffisamment de produits ont été accumulés pendant la première phase,
- *la phase de plateau* qui correspond à une modification du milieu réactionnel ne permettant plus l'accroissement exponentiel.

Au cours de la phase exponentielle, le nombre de molécules amplifiées est représenté par l'équation :

$$N = N_0 * (E)^n \quad [1]$$

Où N est la quantité de molécules amplifiées obtenues au cycle n,  $N_0$  représente la quantité initiale d'ADN cible et E est l'efficacité d'amplification de la PCR. L'efficacité peut être égale à 2 quand 100% des cibles sont reproduites à chaque cycle.

Afin de recueillir des données quantitatives, chacun des échantillons doit être analysé dans sa phase exponentielle d'amplification. La PCR quantitative fait donc le suivi de la fluorescence émise pendant la réaction.

### III.1.2 – LA TECHNOLOGIE DE DETECTION BASEE SUR LE SYBR GREEN

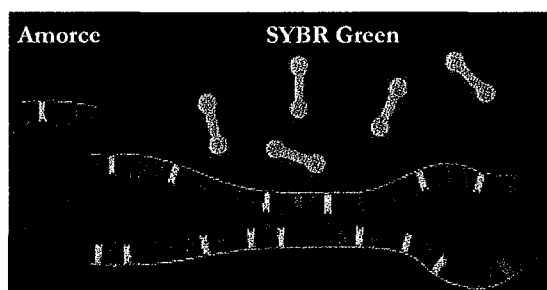
La technologie de la PCR quantitative en temps réel est basée sur la détection et la quantification d'un indicateur fluorescent de la production d'amplicons durant chaque cycle ; l'augmentation du signal fluorescent étant directement proportionnelle à la quantité d'amplicons générés durant la réaction de PCR.

Deux principes généraux sont utilisés pour la détection quantitative des amplicons : les agents se liant à l'ADN double brin (comme par exemple, le SYBR Green) et les sondes fluorescentes.

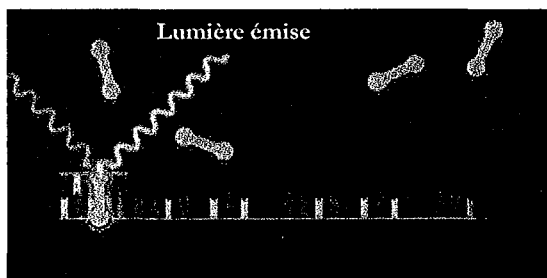
La technologie basée sur le SYBR Green, utilisée dans ce travail, présente l'avantage de pouvoir être appliquée à toute amplification puisque sa spécificité repose entièrement sur les amorces, contrairement aux sondes fluorescentes.

Le SYBR Green est donc un fluorophore capable de se lier à l'ADN double brin (Skeidsvoll and Ueland, 1995). En solution, le SYBR Green n'est que peu fluorescent lorsqu'il est sous sa forme libre, alors qu'il est très fluorescent lorsqu'il est lié à l'ADN. Il est très stable : au bout de 30 cycles d'amplification, seulement 6% de son activité est perdue. Le microspectrofluorimètre couplé à l'appareil permet de lire cette fluorescence.

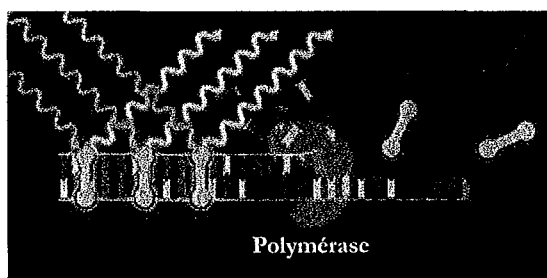
Le principe de la quantification des amplicons par la technologie basée sur le SYBR Green est exposé sur la Figure MM-17.



Au début de l'amplification, le mélange réactionnel contient de l'ADN dénaturé, les amorces et le fluorophore non lié.



Après le couplage des amorces, le fluorophore se lie au double brin. La liaison à l'ADN se traduit par une augmentation de la fluorescence (①).



Pendant l'étape d'élongation, le nombre de molécules de fluorophore lié à l'ADN synthétisé augmente ce qui se traduit par une augmentation de la fluorescence. La fluorescence est mesurée à la fin de l'étape d'élongation de chaque cycle (②).

Figure MM-17 : Principe de la quantification des amplicons par la technologie basée sur le SYBR Green.

Les symboles ① et ② renvoient aux courbes d'amplification présentées en Figure MM-18.



En observant la quantité de fluorescence émise à chaque cycle, il devient ainsi possible d'accéder à la cinétique d'amplification et notamment de suivre la réaction de PCR durant sa phase exponentielle où la première augmentation significative de la quantité d'amplicons est en corrélation directe avec la quantité initiale de matrice cible.

La Figure MM-18 représente des courbes d'amplification pour différentes quantités initiales de cibles.

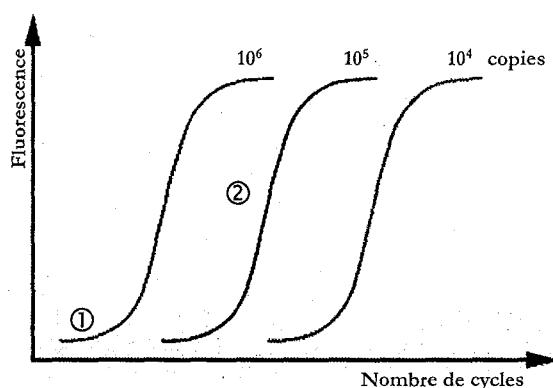


Figure MM-18 : Cinétiques de PCR en temps réel pour différentes quantités de molécules cibles au départ.

Les symboles ① et ② sont expliqués en Figure MM-17.

### III.1.3 - LA NOTION DE CYCLE SEUIL OU CT (THRESHOLD CYCLE)

Comme le montre la Figure MM-18, le démarrage de la phase exponentielle est plus ou moins tardif en fonction de la quantité d'ADN présente initialement. Le Ct, abréviation du terme anglais « threshold cycle », correspond au cycle de démarrage de la phase exponentielle d'amplification. Graphiquement (Figure MM-19), le Ct correspond au point d'intersection de la courbe d'amplification avec la valeur du seuil représentée par le « threshold » (ligne rouge sur la Figure MM-19).

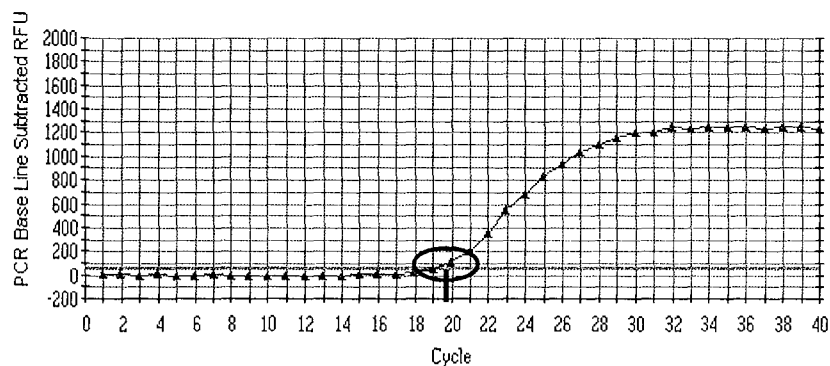


Figure MM-19 : Représentation graphique du cycle seuil ou Ct.

La valeur du Ct peut être traduite en un résultat quantitatif, en la comparant avec les valeurs du Ct générées avec des matrices d'ADN de quantités connues (voir paragraphe II.1.4.1).

### III.1.4 – EFFICACITE ET SPECIFICITE DE LA REACTION

Plusieurs contraintes, reposant sur le principe même de la technique, sont nécessaires à la réalisation d'une expérience de PCR quantitative. En effet les efficacités de PCR sur le gène de référence et sur le gène cible doivent être comparables. Pour cela, la taille des amplicons ainsi que leurs conditions d'amplification doivent être similaires. Les oligonucléotides choisis pour chaque réaction de PCR sont donc définis selon plusieurs critères :

- Leur taux en bases G et C
- Leur Tm
- La taille de l'amplicon
- La présence de moins de 2 bases G ou C dans les 5 dernières bases en 3' de l'oligonucléotide

Les oligonucléotides sont ensuite testés afin de tester la spécificité et l'efficacité de la PCR.

#### III.1.4.1 – EFFICACITE DE LA PCR ET COURBE STANDARD

Afin de déterminer l'efficacité de chaque réaction d'amplification, une courbe standard est réalisée. Pour cela, une gamme de dilutions d'une matrice d'ADN est amplifiée à l'aide d'un même couple d'oligonucléotides. La courbe standard est obtenue en représentant les différentes valeurs de Ct obtenues en fonction du logarithme de la quantité initiale de matrice (la quantité d'ADN correspondant à la dilution la plus élevée étant arbitrairement fixée à 1, Figure MM-20).

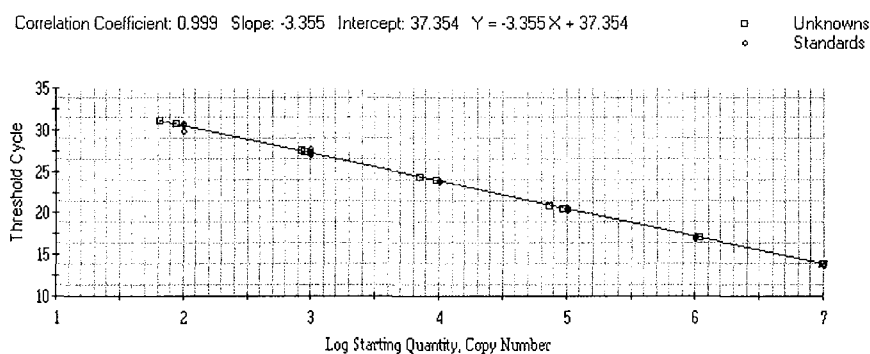


Figure MM-20 : Droite standard du Ct en fonction du logarithme de la quantité initiale de matrice d'ADN.

L'équation de cette droite (équation [5], voir ci-après) permet d'accéder à l'efficacité de la réaction de PCR ainsi qu'à la reproductibilité de l'amplification grâce au coefficient de corrélation.

En effet, en remplaçant la valeur du cycle  $n$  par la valeur du  $C_t$  dans l'équation représentant le nombre de molécules amplifiées au cours de la phase exponentielle (équation [1], voir paragraphe III-1.1), nous pouvons alors écrire :

$$N = N_0 * (E)^{C_t} \quad [2]$$

où  $N$  est la quantité de molécules amplifiées obtenues au  $C_t$ ,  $N_0$  représente la quantité initiale d'ADN cible et  $E$  est l'efficacité d'amplification de la PCR.

Cette équation peut alors être transformée :

$$\log N = \log N_0 + C_t * \log E \quad [3]$$

A partir de l'équation [3], le  $C_t$  peut être déduit sous la forme d'une équation linéaire :

$$C_t = -(1/\log E) * \log N_0 + \log N / \log E \quad [4]$$

Cette équation est alors de la forme :

$$C_t = -a * \log N_0 + b \quad [5]$$

L'équation [5] correspond à l'équation de la droite standard où «  $a$  », la pente de la droite, est l'inverse négative du logarithme de l'efficacité.

L'efficacité de la réaction de PCR est donc donnée par l'équation suivante :

$$E = (10^{1/a} - 1) * 100 \quad [6]$$

Ainsi, dans l'exemple de droite standard donné en Figure MM-20, l'efficacité de la réaction de PCR peut être calculée à partir de la valeur de pente de la droite :

$$E = (10^{1/-3,355} - 1) * 100 = 98,7\%$$

#### III.1.4.2 – SPECIFICITE DE LA PCR ET COURBE DE FUSION

Afin de déterminer la spécificité de la réaction, une courbe de fusion est réalisée à la fin de l'amplification.

En effet, chaque produit d'ADN double brin synthétisé a une température de fusion ( $T_m$ ) spécifique, définie comme étant la température à partir de laquelle 50 % de l'ADN est sous forme double brin et 50 % sous forme simple brin. Après le dernier cycle de PCR, la température est rapidement élevée à 95°C pour dénaturer l'ADN double brin, puis elle est abaissée la température de couplage ce qui provoque la reformation d'ADN double brin. La température est ensuite élevée lentement à 95°C. La fluorescence est lue en continu pendant cette remontée (Figure MM-21 A.) :

- le SYBR Green est fluorescent tant qu'il est lié à l'ADN double brin,
- quand la température augmente, l'ADN double brin se dissocie, ce qui entraîne la libération du SYBR Green dans le milieu et une diminution progressive de la fluorescence,

□ lorsque 50 % de l'ADN double brin sont dissociés, la fluorescence chute brutalement, c'est à cette température que correspond à la température de fusion du produit synthétisé.

La température de fusion est obtenue en traçant la dérivée première négative de la fluorescence en fonction de la température : elle correspond au maximum de la courbe (Figure MM-21 B.).

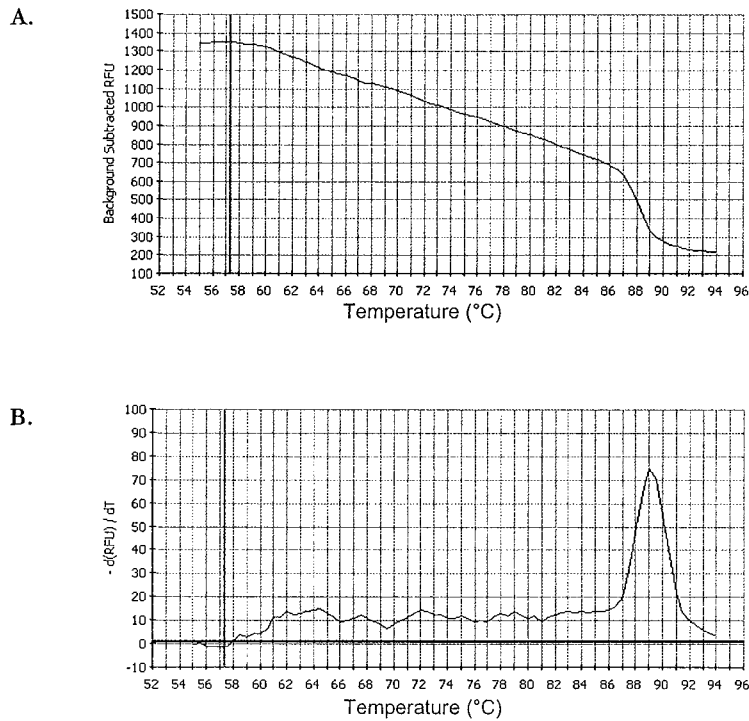


Figure MM-21 : Courbe de fusion déterminant la spécificité de la réaction de PCR.

A. Graphique représentant la quantité de fluorescence mesurée en fonction de la température,  
 B. Graphique représentant la fonction dérivée de A. permettant d'accéder à la température de fusion spécifique de l'amplicon (ici environ 89°C).

Ainsi, la spécificité de la réaction est révélée par la présence d'un seul pic lors de cette analyse. De plus, les dimères d'amorces pouvant se former au cours de la réaction sont également révélés par cette technique de courbe de fusion sous forme d'un pic à une température inférieure et d'amplitude également plus faible que celle de l'amplicon lui-même.

Ainsi, les couples d'oligonucléotides retenus pour les expériences ultérieures seront ceux dont l'analyse par courbe de fusion ne révèle qu'un seul pic et donnant une efficacité de PCR comprise entre 80 et 120%.

### III.2 ☞ UN EXEMPLE D'ANALYSE DE RESULTATS DE PCR QUANTITATIVE

Les expériences de PCR quantitative sont réalisées dans le « thermocycleur » iCycler™ iQ de chez Bio-Rad (Figure MM-22) et les résultats analysés à l'aide du logiciel fourni avec l'appareil (iCycler™ iQ Optical system software Version 3.0a).

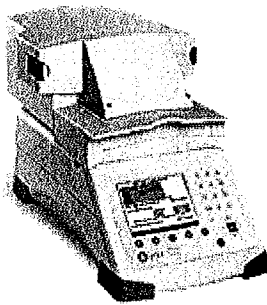


Figure MM-22 : Photographie de l'appareillage iCycler™ iQ (Bio-Rad).

La technologie de PCR quantitative passe par un traitement mathématique des données, basé sur le calcul du  $\Delta\Delta Ct$  (réf).

Cette méthode de calcul postule que les efficacités des réactions de PCR sont de 100%.

Un exemple d'analyse de résultats de PCR quantitative à l'aide de la méthode du  $\Delta\Delta Ct$  va être exposé.

L'objectif est de quantifier la transcription d'un gène cible (par exemple, un gène *has*) dans deux conditions expérimentales distinctes (par exemple, deux phases de croissance). La technique de PCR quantitative nécessite de disposer d'un « gène de référence », dont l'expression est constante quelles que soient les conditions testées.

Le gène de référence utilisé lors de nos expériences est le gène *brdB*, codant le facteur  $\sigma$  principal de *S. coelicolor*. Ce gène est effectivement fréquemment utilisé comme référence dans les expériences classiques de quantification de transcrits par Northern Blot (comparaison schématique des résultats obtenus par cette technique avec ceux obtenus par PCR quantitative en Figure MM-23) et présente une expression relativement au cours de la croissance et dans les diverses conditions testées dans de ce travail.

Ainsi, après avoir vérifié la spécificité des réactions de PCR, des courbes standards sont alors réalisées pour chaque couple d'amorces, afin de vérifier que les efficacités sont effectivement de l'ordre de 100% et surtout comparables entre le gène cible et le gène de référence. Les expériences ultérieures ne nécessitent pas la réalisation systématique de ces courbes standards, puisque seules les valeurs des Ct sont utilisées dans la méthode de calcul du  $\Delta\Delta Ct$ .

Après avoir réalisés ces mises au points, des expériences de RT-PCR quantitative peuvent alors être réalisées à partir de d'échantillons d'ARN extraits dans les conditions à tester (dans notre exemple, les ARN auront été extraits à partir de cultures en phases exponentielle et de transition de croissance).

Chaque échantillon d'ARN sera analysé (en duplicats) par RT-PCR quantitative à l'aide des deux couples d'amorces, spécifiques du gène cible et gène de référence.

La Figure MM-23 présente les courbes d'amplification et les valeurs de Ct ainsi obtenues.

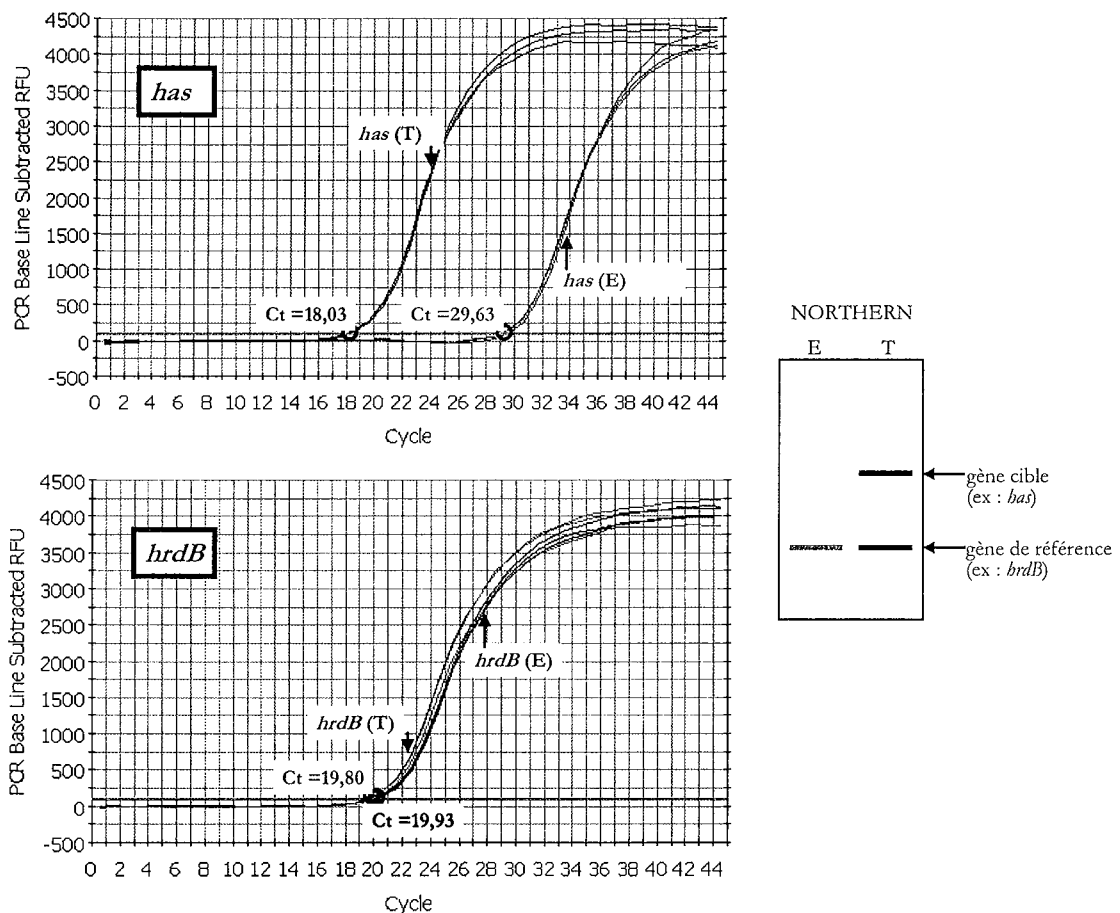


Figure MM-23 : Quantification d'un transcrite d'un gène cible en comparaison à un gène de référence dans deux conditions expérimentales distinctes par les techniques de Northern Blot et de RT-PCR quantitative.

A gauche : Courbes d'amplification d'un gène contrôlé (ex : *has*) et du gène de référence (ex : *hrdB*) obtenues par RT-PCR quantitative, à partir d'ARN extraits dans deux conditions expérimentales distinctes (ex : phase exponentielle et phase de transition). Les valeurs de Ct indiquées sont les moyennes obtenues à partir des duplicats.

A droite : Schématisation des résultats d'un Northern Blot réalisé à partir des mêmes ARN.

Dans notre exemple, nous disposons donc, à l'issue de cette expérience, de 4 valeurs de Ct (Figure MM-23), permettant de déterminer le facteur d'induction de la transcription du gène *bas* entre la phase exponentielle de croissance et la phase de transition par le calcul du  $\Delta\Delta Ct$ .

Dans un premier temps, la différence de Ct (ou  $\Delta Ct$ ) entre le gène cible et le gène de référence doit être calculée pour chaque condition expérimentale :

Phase exponentielle :

$$\Delta Ct_E = Ct \text{ Cible } (bas) - Ct \text{ Référence } (brdB) = 29,63 - 19,63 = 9,70$$

Phase de transition :

$$\Delta Ct_T = Ct \text{ Cible } (bas) - Ct \text{ Référence } (brdB) = 18,03 - 19,80 = -1,7$$

Le calcul du  $\Delta\Delta Ct$  consiste à comparer ces différences de Ct. Ainsi, une condition expérimentale doit être considérée comme référence, en générale la condition de croissance normale ou la phase de croissance où le gène cible présente le plus faible niveau de transcription. Dans ce travail, les facteurs d'induction ont été déterminés par rapport à la phase exponentielle de croissance (dans les expériences visant à quantifier l'expression des gènes *bas* au cours de la croissance) et par rapport à la croissance en milieu HT (pour les expériences visant à caractériser l'induction transcriptionnelle des gènes *bas* dans des conditions de stress).

Dans notre exemple, la phase exponentielle de croissance sera la condition expérimentale de référence. Les valeurs des  $\Delta\Delta Ct$  sont donc calculées de la manière suivante :

Phase exponentielle :

$$\Delta\Delta Ct_E = \Delta Ct_E - \Delta Ct_E = 0$$

Phase de transition :

$$\Delta\Delta Ct_T = \Delta Ct_T - \Delta Ct_E = 9,70 - (-1,7) = 11,40$$

La dernière étape du calcul consiste à replacer cette valeur du  $\Delta\Delta Ct$  dans l'équation [2] (voir III-1), tout en considérant également que la quantité initiale de matrice d'ADN ( $N_0$ ) est égale à 1 puisque nous cherchons à déterminer des facteurs d'induction ( $N$ ). Ainsi, nous pouvons écrire :

$$N = 1 * (E)^{\Delta\Delta Ct} \quad [7]$$

Le calcul du  $\Delta\Delta Ct$  étant basé sur le postulat selon lequel les efficacités de PCR sont de 100%, nous pouvons alors transformer cette équation :

$$N = 2^{\Delta\Delta Ct} \quad [8]$$

Ainsi, dans notre exemple, nous pouvons calculer le facteur d'induction de la transcription du gène *bas* :

Phase exponentielle :

$$N_E = 2^{\Delta\Delta_{GtE}} = 2^0 = 1$$

Phase de transition :

$$N_T = 2^{\Delta\Delta_{GtT}} = 2^{11,40} = 2702$$

Cette expérience aurait donc permis de montrer que le gène *bas* était 2702 fois plus transcrit en phase de transition qu'en phase exponentielle de croissance.

L'ensemble des résultats de RT-PCR quantitative obtenus au cours de ce travail ont été obtenus par cette méthode et sont présentés sous forme des facteurs d'induction (N, voir notamment Tableau R-11).



## ∞ RESULTATS ∞

**A - PREMIERE PARTIE :**

**DYNAMIQUE DE LA DUPLICATION CHROMOSOMIQUE CHEZ**

***STREPTOMYCES AMBOFACIENS* \_\_\_\_\_ 119**

**- CHAPITRE A-I -**

**MISE EN EVIDENCE DE FUSIONS DE CHROMOSOMES PARTIELLEMENT DELETES**

<b>A-I.1</b> $\propto$ Structure chromosomique fusionnée des mutants <i>S. ambofaciens</i> NSA27 et NSA65 _____	<b>121</b>
A-I.1.1 – Le mutant <i>S. ambofaciens</i> NSA27 _____	121
A-I.1.2 – Le mutant <i>S. ambofaciens</i> NSA65 _____	125
<b>A-I.2</b> $\propto$ Analyse des séquences impliquées dans le réarrangement chez <i>S. ambofaciens</i> NSA27 _____	<b>128</b>

**- CHAPITRE A-II -**

**LA FUSION CHROMOSOMIQUE : UN ETAT MUTATEUR ?**

<b>A-II.1</b> $\propto$ Fusion de chromosomes et instabilité chromosomique _____	<b>131</b>
A-II.1.1 – Instabilité chromosomique des mutants <i>S. ambofaciens</i> NSA27 et NSA65 _____	131
A-II.1.2 - Un cycle d'instabilité chromosomique _____	133
<b>A-II.2</b> $\propto$ Fusion de chromosomes et instabilité phénotypique _____	<b>135</b>
A-II.2.1 – Instabilité phénotypique des mutants <i>S. ambofaciens</i> NSA27 et NSA65 et de leurs descendants _____	135
A-II.2.2 – Corrélation entre état fusionné et défaut de sporulation _____	136
A-II.2.2.1 – Analyse du contenu en ADN des spores par coloration au DAPI _____	136
A-II.2.2.2 – Analyse de l'ultrastructure des spores par microscopie électronique _____	137
A-II.2.2.3 – Analyse du taux de germination _____	139

**B - SECONDE PARTIE :**

**UN EVENEMENT DE DUPLICATION GENIQUE DANS LES REGIONS  
TERMINALES DU CHROMOSOME DE  
*STREPTOMYCES AMBOFACIENS* \_\_\_\_\_ 140**

**- CHAPITRE B-I -**

**CARACTERISATION DES GENES *HAS***

<b>B-I.1</b> ☞ Les gènes <i>hasR</i> et <i>hasL</i> chez la souche sauvage <i>Streptomyces ambofaciens</i> DSM40697 _____	142
B-I.1.1 – Les facteurs $\sigma^{\text{Has}}$ _____	142
B-I.1.2 – L’environnement génétique des gènes <i>has</i> _____	146
<b>B-I.2</b> ☞ Les gènes chimériques <i>has135</i> et <i>has855</i> chez les mutants <i>Streptomyces ambofaciens</i> NSAH135 et NSA855 _____	151

**- CHAPITRE B-II -**

**LA DUPLICATION DES GENES *HAS* CHEZ *STREPTOMYCES AMBOFACIENS* EST RECENTE**

<b>B-II.1</b> ☞ Conservation de l’architecture de la région « core » du chromosome de <i>S. ambofaciens</i> _____	154
B-II.1.1 – Détermination des séquences des ITS de <i>S. ambofaciens</i> _____	154
B-II.1.2 – Mise en évidence de blocs constants et variables dans les ITS de <i>Streptomyces</i> _____	155
B-II.1.3 – Comparaison des séquences ITS _____	156
B-II.1.4 – Prédiction des structures secondaires du repliement des séquences ITS _____	157
<b>B-II.2</b> ☞ Conservation de la duplication des gènes <i>has</i> chez les trois souches de <i>Streptomyces ambofaciens</i> _____	159
<b>B-II.3</b> ☞ Les gènes <i>has</i> au sein du genre <i>Streptomyces</i> _____	163

**- CHAPITRE B-III -**

**ANALYSE TRANSCRIPTIONNELLE DES GENES DUPLIQUES *HASR* ET *HASL***

<b>B-III.1</b>	<b>œ Analyse qualitative de la transcription des gènes <i>has</i> par RT-PCR</b>	<b>167</b>
B-III.1.1	– Expression des gènes <i>hasR</i> et <i>hasL</i> chez la souche sauvage <i>S. ambfaciens</i> DSM40697	167
B-III.1.2	– Expression des gènes <i>hasR</i> et <i>hasL</i> et des chimères <i>has855</i> et <i>has135</i> au cours de la croissance	169
<b>B-III.2</b>	<b>œ Identification des transcrits <i>has</i> et analyse de la régulation de leur expression</b>	<b>173</b>
<hr/>		
B-III.2.1	– Identification des sites d'initiation de la transcription et des promoteurs des gènes <i>has</i>	173
B-III.2.1.1	– Identification des +1 des gènes <i>has</i> par une approche d'extension d'amorces	173
B-III.2.1.2	– Identification des sites +1 des gènes <i>has</i> par une approche de cartographie à la nucléase S1 et analyse des séquences promotrices	175
B-III.2.1.2.1	- Les transcrits du gène <i>hasR</i>	175
B-III.2.1.2.1.1	– Le transcrit initié au site <i>hasRp1</i> et son promoteur de type $\sigma^B$	175
B-III.2.1.2.1.2	– Le transcrit initié au site <i>hasRp2</i> et son promoteur de type $\sigma^B$	177
B-III.2.1.2.1.3	– Les promoteurs putatifs du gène <i>hasR</i> reconnus par <i>WhiG</i>	178
B-III.2.1.2.2	- Les transcrits du gène <i>hasL</i>	180
B-III.2.1.2.2.1	– Le transcrit initié au site <i>hasLp1</i> et son promoteur de type $\sigma^B$	181
B-III.2.1.2.2.2	– Le transcrit initié au site <i>hasLp2</i> et son promoteur de type $\sigma^{HrdB}$	182
B-III.2.2	– Régulation de l'expression des transcrits <i>has</i> au cours de la croissance	183
B-III.2.2.1	– Identification de la régulation de la transcription des gènes <i>has</i> au cours de la croissance par cartographie à la nucléase S1	183
B-III.2.2.2	– Quantification des niveaux d'induction de la transcription des gènes <i>has</i> au cours de la croissance par PCR quantitative	184
B-III.2.2.2.1	– Principes	184
B-III.2.2.2.1.1	– Principe de la quantification de chaque transcrit <i>has</i>	184
B-III.2.2.2.1.2	– Traitement mathématique des données	186
B-III.2.2.2.2	– Quantification de la transcription du gène <i>hasR</i>	187
B-III.2.2.2.2.1	– Evolution de la quantité de chaque transcrit du gène <i>hasR</i> au cours de la croissance	187
B-III.2.2.2.2.2	– Abondance relative des transcrits initiés aux sites <i>hasRp1</i> et <i>hasRp2</i>	188
B-III.2.2.2.3	– Quantification de la transcription du gène <i>hasL</i>	188
B-III.2.2.2.3.1	– Evolution de la quantité de chaque transcrit du gène <i>hasL</i> au cours de la croissance	188
B-III.2.2.2.3.2	– Abondance relative des transcrits initiés aux sites <i>hasLp1</i> et <i>hasLp2</i>	189
B-III.2.2.2.4	– Expression relative des gènes <i>hasR</i> et <i>hasL</i> au cours de la croissance	191
<b>B-III.3</b>	<b>œ Analyse de l'expression spatio-temporelle des promoteurs <i>has</i></b>	<b>192</b>
III.3.1	– Utilisation du gène rapporteur codant la résistance à la kanamycine	192

B-III.3.1.1 - Clonage des régions promotrices des gènes <i>hasR</i> et <i>hasL</i> _____	192
B-III.3.1.2 - Cartographie de restriction du plasmide pNSL9 _____	194
B-III.3.1.2.1 - Cartographie de restriction BstBI du plasmide pNSL9 _____	194
B-III.3.1.2.2 - Cartographie de restriction ApaI du vecteur recombiné pNSL9 _____	195
B-III.3.1.3 - Détection d'une force promotrice en amont de <i>hasL</i> _____	196
B-III.3.1.3.1 - Expression de la résistance à la kanamycine chez <i>S. lividans</i> NSL9 en croissance sur milieu solide _____	196
III.3.1.3.2 - Expression de la résistance à la kanamycine chez <i>S. lividans</i> NSL9 au cours de la croissance en milieu liquide _____	197
B-III.3.2 – Utilisation du gène rapporteur codant L'EGFP _____	198
B-III.3.2.1 - Clonage des régions promotrices des gènes <i>hasL</i> et <i>hasR</i> dans le vecteur pIJ8630 _____	198
B-III.3.2.2 - Cartographie de restriction des plasmides pIJR et pIJL _____	200
B-III.3.2.2.1 – Les vecteurs recombinants pIJL _____	200
B-III.3.2.2.2 – Les vecteurs recombinants pIJR _____	202
B-III.3.2.3 - Analyse de la fluorescence chez <i>S. ambofaciens</i> _____	204

#### - CHAPITRE B-IV -

#### CONSTRUCTION DE MUTANTS INTERROMPUS POUR LE(S) GENE(S) *HAS*

<b>B-IV.1</b> ☞ Stratégie de remplacement d'une ORF au locus chromosomique chez <i>Streptomyces</i> _____	207
B-IV.1.1 – Choix des amorces et fabrication du produit de PCR (étape 1) _____	209
B-IV.1.2 – Construction de l'allèle modifié sur le cosmide recombinant (étape 2) _____	209
B-IV.1.3 – Transfert conjugatif du cosmide modifié d' <i>E. coli</i> vers <i>Streptomyces</i> et sélection des recombinants _____	210
<b>B-IV.2</b> ☞ Construction des simples mutants $\Delta hasR$ et $\Delta hasL$ _____	211
B-IV.2.1 – Introduction des cosmides 14C4 et 25E1 dans la souche <i>E. coli</i> BW25113/pIJ790 _____	211
B-IV.2.2 – Amplification des cassettes Apra _____	212
B-IV.2.3 – Remplacement des ORF <i>hasL</i> et <i>hasR</i> par la cassette Apra dans les cosmides recombinants _____	212
B-IV.2.3.1 – Electroporation et sélection _____	212
B-IV.2.3.2 – Analyse des transformants _____	213
B-IV.2.4 – Remplacement des ORF <i>has</i> aux loci chromosomiques chez <i>S. ambofaciens</i> DSM40697 _____	214
B-IV.2.4.1 – Introduction des cosmides modifiés dans la souche <i>E. coli</i> ET12567/pUZ8002 _____	214

B-IV.2.4.2 – Conjugaisons intergénériques entre <i>E. coli</i> ET12567/pUZ8002/14C4 $\Delta$ <i>hasL</i> ::Apra ou 25E1 $\Delta$ <i>hasR</i> ::Apra et <i>S. ambofaciens</i> DSM40697	215
B-IV.2.4.3 – Mise en évidence de l'échange allélique aux loci <i>has</i> chromosomiques	215
B-IV.2.4.3.1 – Vérification par hybridation	215
B-IV.2.4.3.2 – Vérification par PCR	217
B-IV.2.4.4 – Vérification de l'intégrité de la structure chromosomique chez les simples mutants <i>S. ambofaciens</i> $\Delta$ <i>has</i> ::Apra	219
<b>B-IV.3 <math>\otimes</math> Construction de doubles mutants <math>\Delta</math><i>hasR</i> <math>\Delta</math><i>hasL</i></b>	<b>220</b>
B-IV.3.1 – Amplification des cassettes Spec et remplacement dirigé des loci <i>has</i> sur les cosmides recombinants 14C4 et 25E1	220
B-IV.3.2 – Remplacement du locus <i>has</i> sauvage par la cassette Spec chez les simples mutants $\Delta$ <i>hasL</i> ::Apra et $\Delta$ <i>hasR</i> ::Apra	222
B-IV.3.3 – Mise en évidence du remplacement des deux ORF <i>has</i> aux loci chromosomiques	222
B-IV.3.3.1 – Vérification par hybridation	222
B-IV.3.3.2 – Vérification par PCR	223
IV.3.4 – Intégrité des régions terminales du chromosome des doubles mutants	224
<b>B-IV.4 <math>\otimes</math> Construction de complémentants des simples mutants <math>\Delta</math><i>has</i></b>	<b>226</b>
B-IV.4.1 – Stratégie	226
B-IV.4.2 – Clonage des régions codantes des gènes <i>hasL</i> et <i>hasR</i> dans le vecteur pSET $\Omega$	227
B-IV.4.3 – Introduction chez <i>streptomyces</i>	227

- CHAPITRE B-V -

ANALYSE PHENOTYPIQUE DES MUTANTS  $\Delta$ *HAS*

<b>B-V.1 <math>\otimes</math> Les facteurs <math>\sigma^{\text{Has}}</math> et la différenciation cellulaire</b>	<b>230</b>
B-V.1.1 – Croissance sur milieux gélosés	230
B-V.1.2 – Activité antibactérienne	232
<b>B-V.2 <math>\otimes</math> Les facteurs <math>\sigma^{\text{Has}}</math> et la réponse aux stress</b>	<b>233</b>
B-V.2.1 – Recherche de phénotypes associés aux stress	233
B-V.2.1.1 – Influence de la température sur le phénotype colonial	233
B-V.2.1.2 – Stress acide-alcalin	234
B-V.2.1.3 – Stress osmotique	234
B-V.2.1.4 – Stress UV	235
B-V.2.1.5 – Stress oxydant	236
B-V.2.2 – Analyse transcriptionnelle des gènes <i>has</i> en conditions de stress	238



Les réarrangements constituent un moteur pour l'évolution de la structure chromosomique. Chez *Streptomyces ambofaciens* existe une instabilité génétique qui a été corrélée à l'apparition de réarrangements de grande ampleur.

La bactérie *Streptomyces ambofaciens* peut être considérée comme un modèle d'étude des phénomènes d'évolution des génomes par duplication du matériel génétique, et ce aussi bien en termes de dynamique du phénomène qu'en termes évolutifs.

□ La dynamique du phénomène de duplication peut en effet être abordée grâce à l'analyse de la descendance de mutants issus de l'instabilité génétique de *S. ambofaciens* au caractère mutateur, puisque générant une très grande variabilité chromosomique. En effet, la dynamique des phénomènes d'instabilité génétique chez les *Streptomyces* souligne la plasticité spectaculaire du chromosome, notamment au niveau des régions subterminales du chromosome linéaire, puisque notamment chez *S. ambofaciens*, la délétion de régions subterminales incluant le(s) télomère(s) bactérien(s) induit la formation de réarrangements génomiques importants. L'analyse de la structure du chromosome de souches mutantes portant une délétion terminale incluant le télomère et présentant une duplication de la quasi totalité de leur génome ainsi que la caractérisation de la variabilité chromosomique générée dans la descendance de ces souches sont présentés dans la première partie de ce manuscrit.

□ Les réarrangements de l'instabilité génétique, confinés aux extrémités du chromosome linéaire (8 mégabases), peuvent toucher jusqu'à 25 % du génome, soit environ 2000 kb. Bien que non essentielle en conditions de laboratoire, cette région contient de nombreux gènes. L'analyse du polymorphisme des régions terminales chez les *Streptomyces* a révélé une compartimentation de l'information génétique. Le séquençage complet du génome de *S. coelicolor* A3(2) a confirmé cette notion en attribuant aux gènes localisés dans ces régions des fonctions impliquées dans l'expression du métabolisme secondaire et dans la réponse aux changements environnementaux. L'impact des réarrangements chromosomiques sur l'expression génique peut donc être abordé chez les *Streptomyces* en étudiant l'expression des gènes localisés dans les régions terminales du chromosome.

La caractérisation de deux souches mutantes de *Streptomyces ambofaciens* NSAH135 et NSA855 présentant des translocations chromosomiques (Fischer *et al.*, 1998b) a permis de mettre en évidence la duplication d'une ORF codant un facteur sigma alternatif (appelées *has* pour "homologous to alternative sigma factor", et R et L pour leur localisation respective sur le bras chromosomique droit et gauche). Le réarrangement chromosomique présent chez les mutants pourrait consister en un événement de recombinaison homologue ectopique entre les séquences *hasR* et *hasL* (Figure R-1). Outre la délétion et la duplication de nombreux gènes, la conséquence de ces échanges d'extrémités chromosomiques est la formation d'ORF chimériques au point de recombinaison. Les gènes *hasR* et *hasL*, localisés dans les régions terminales du chromosome de *S. ambofaciens*, constituent donc un modèle expérimental de choix pour aborder l'impact de tels événements de duplication sur l'expression génique chez *S. ambofaciens*.

Après avoir démontré une certaine conservation au niveau intraspécifique de l'organisation du chromosome des *Streptomyces* (via l'analyse du polymorphisme des séquences ITS chez *S. ambofaciens* ; paragraphe II.1 de la seconde partie), le cas de la duplication des gènes *has* chez *S. ambofaciens*, comparativement aux autres espèces de *Streptomyces* ne présentant pas cette



duplication, permet d'aborder la problématique de la fixation de tels événements de duplication génique au cours de l'évolution (seconde partie).

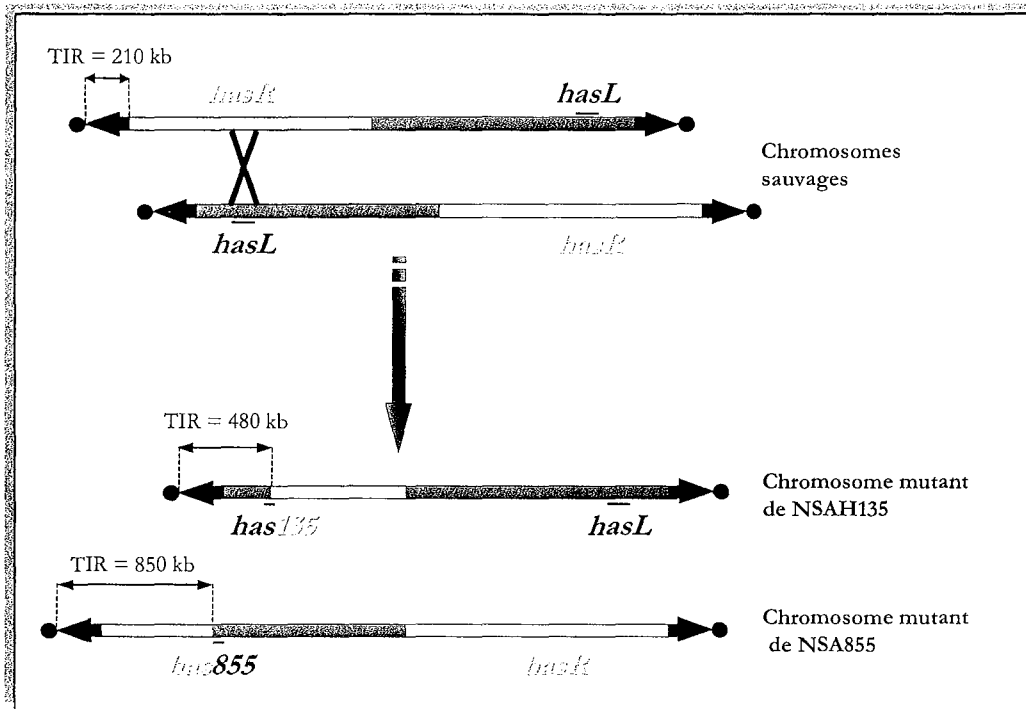


Figure R-1 : Représentation de la structure chromosomique des souches mutantes *S. ambifaciens* NSAH135 et NSA855.

Les bras chromosomiques gauche et droit sont arbitrairement séparés par l'origine de réplication principale *oriC*. Les protéines terminales du chromosome sont représentées par des sphères noires.

Le réarrangement chromosomique présent chez les souches mutantes est ici expliqué selon l'hypothèse d'une recombinaison homologue entre les gènes *hasR* et *hasL*.

Les résultats présentés dans ce chapitre ont fait l'objet de trois publications dont une est actuellement soumise à *Journal of Bacteriology*.

Les travaux publiés apparaîtront sous forme résumée dans la première partie et dans le chapitre II-1 de la seconde partie de ces résultats.

□ **End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens***

Thomas Wenner, Virginie Roth, Gilles Fischer, Céline Fourier, Bertrand Aigle ,  
Bernard Decaris and Pierre Leblond  
*Molecular Microbiology, sous presse*

□ **Intragenomic and intraspecific polymorphism of the 16S-23S rDNA internally transcribed sequences of *Streptomyces ambofaciens***

Thomas Wenner, Virginie Roth, Bernard Decaris and Pierre Leblond  
*Microbiology, Vol.148, pp 633-642*

La seconde partie traite de façon détaillée de résultats qui font l'objet d'un manuscrit en préparation pour *Journal of Bacteriology*.

□ **Differential and crossed transcriptional expression of duplicated genes which encode alternative sigma factors in *Streptomyces ambofaciens***

Virginie Roth, Bertrand Aigle, Robert Bunet, Thomas Wenner, Céline Fourier,  
Bernard Decaris, and Pierre Leblond  
*Soumis à Journal of Bacteriology*

A - PREMIÈRE PARTIE :

DYNAMIQUE DE LA DUPLICATION  
CHROMOSOMIQUE CHEZ  
*STREPTOMYCES AMBOFACIENS*

**- CHAPITRE A-I -**

**MISE EN EVIDENCE DE FUSIONS DE  
CHROMOSOMES PARTIELLEMENT  
DELETES**

*Les résultats résumés dans ce chapitre sont exposés dans la publication (Wenner et al., 2002) (page 340).*

## **A-I.1 $\otimes$ STRUCTURE CHROMOSOMIQUE FUSIONNÉE DES MUTANTS *S. AMBOFACIENS* NSA27 ET NSA65**

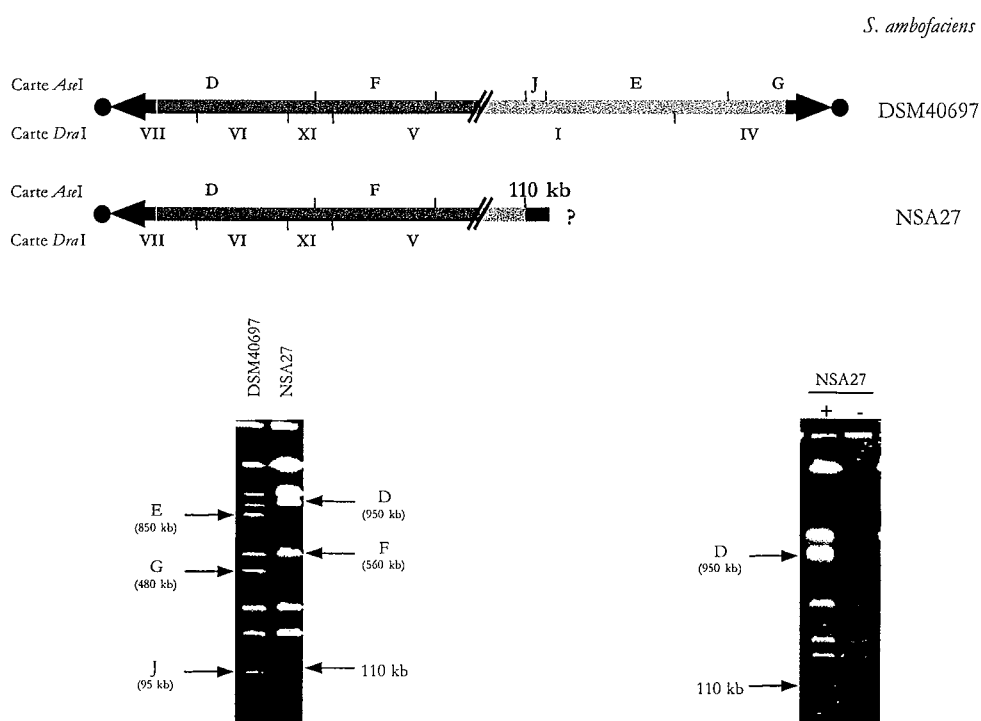
La structure chromosomique de deux souches mutantes *S. ambofaciens* NSA27 et NSA65 (Leblond *et al.*, 1991), isolées dans la descendance de la souche sauvage *S. ambofaciens* DSM40697, a été caractérisée par des expériences de migration en électrophorèse en champs pulsés.

### **A-I.1.1 – LE MUTANT *S. AMBOFACIENS* NSA27**

Le chromosome du mutant NSA27 présente une délétion des fragments *AseI* E, G et J, cartographiés à l'extrémité droite du chromosome sauvage (Figure R-2). De plus, un nouveau fragment d'environ 110 kb a été détecté dans le profil mutant et a été cartographié à l'extrémité droite du chromosome (Figure R-2). Ainsi, la taille de la délétion a pu être estimée à environ 1350 kb.

Afin de tester la présence d'une structure télomérique typique des *Streptomyces* (i.e. une extrémité d'ADN chromosomique liée de façon covalente à une protéine) à l'extrémité du fragment de 110 kb, une comigration de l'ADN du mutant, traité ou non à la protéase, puis digéré par *AseI* a été réalisée. Un retard de migration électrophorétique d'un fragment d'ADN dans l'échantillon non traité à la protéase est révélateur de la présence d'une protéine liée de façon covalente à un fragment d'ADN.

Seul le fragment *AseI*-D du mutant, correspondant à l'extrémité chromosomique gauche intacte, présente un retard de migration électrophorétique (Figure R-2). Ce résultat suggère donc que le fragment de 110 kb ne porte pas de protéines terminales.

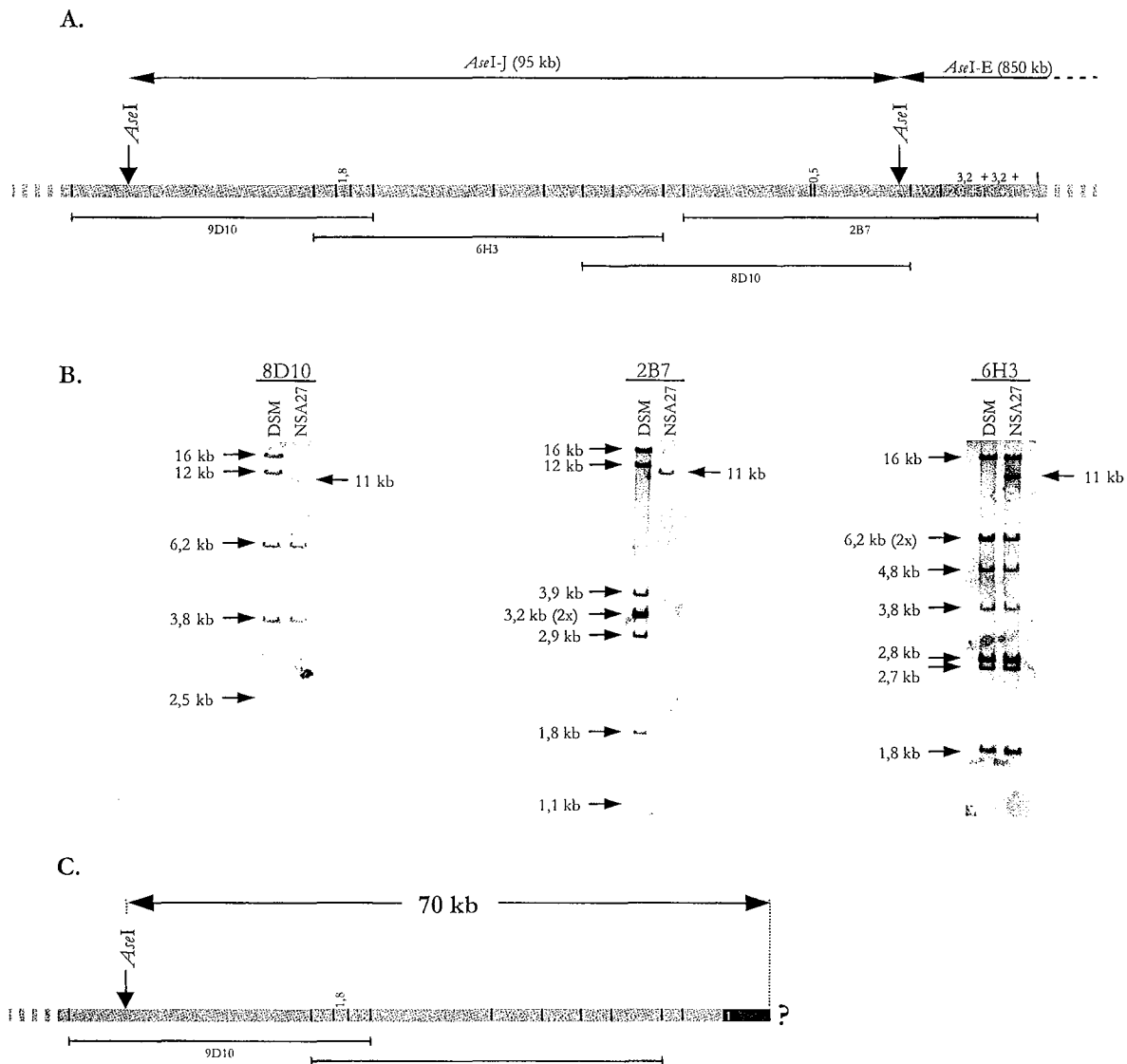


**Figure R-2 : Réarrangement chromosomique dans la souche *S. ambofaciens* NSA27.**

La carte *AseI* de la région instable du chromosome de la souche sauvage est indiquée. Les bras droit et gauche sont représentés en orange et vert respectivement. Les flèches noires représentent les répétitions terminales inversées sauvages et les cercles les protéines terminales. La partie centrale du chromosome est schématisée par deux traits parallèles obliques. Les fragments *AseI* (lettres) et *DraI* (chiffres romains) sont indiqués sur la carte physique (selon Leblond *et al.* (1996)). En bas : à gauche, profil de restriction *AseI* des souches sauvage et mutante ; à droite, profil de restriction *AseI* de l'ADN du mutant NSA27 ayant subi un traitement à la protéase (+) ou n'ayant pas subi ce traitement (-).

Des hybridations successives de l'ADN marqué des cosmides 2B7, 8D10, 6H3 et 9D10, recouvrant le fragment *AseI*-J sauvage, sur des profils de restriction *Bam*HI de l'ADN de NSA27 ont permis de localiser la borne de délétion dans le fragment *Bam*HI de 16 kb porté par les cosmides 8D10 et 2B7 (Figure R-3A).

De plus, il est apparu qu'un fragment de jonction *Bam*HI de 11 kb était produit par le réarrangement ; ce fragment étant révélé par les cosmides 8D10, 2B7 mais aussi 6H3.

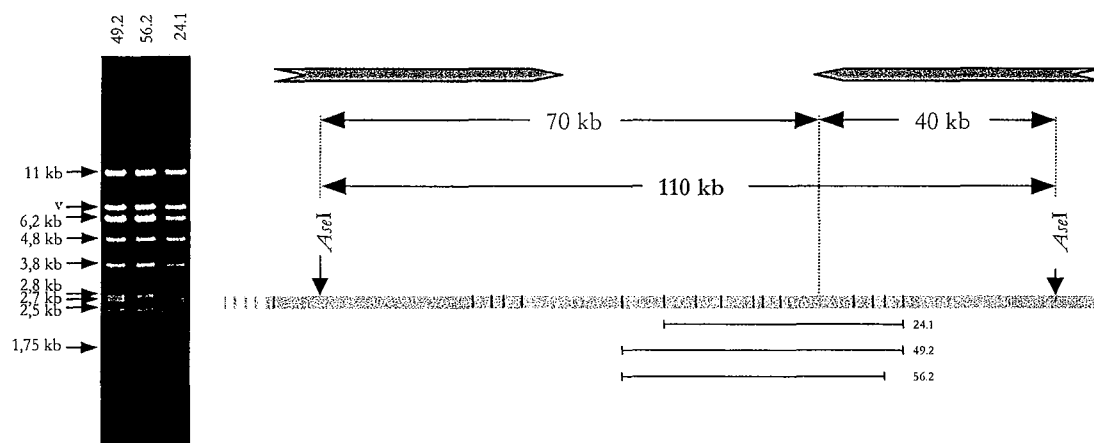


**Figure R-3 :** Caractérisation du réarrangement d'ADN dans la souche *S. ambifaciens* NSA27.  
**A.** Cartographie *Bam*HI du fragment *Ase*I-J et localisation de l'insert des cosmides utilisés lors de ce travail. Les sites *Bam*HI sont représentés par des traits noirs verticaux, et la taille des fragments est indiquée en kilobases.

**B.** Hybridations de l'ADN des cosmides 8D10, 2B7 et 6H3 sur des profils de restriction *Bam*HI des souches sauvage et mutante NSA27.

**C.** Interprétation des résultats des hybridations, montrant la cartographie partielle (70 kb) du fragment de 110 kb mis en évidence chez NSA27. Le point d'interrogation signifie que l'origine de l'ADN additionnel au delà du site *Bam*HI flanquant le fragment de 110 kb n'est pas connu à ce point de raisonnement. Le fragment *Bam*HI de 11 kb est représenté en deux couleurs symbolisant le fait qu'il est à la fois homologue de séquences clonées dans 6H3, 2B7 et 8D10.

La cartographie fine du fragment *Ase*I de 110 kb a alors été entreprise. Ainsi, une banque génomique du mutant NSA27 a été construite et criblée à l'aide de l'ADN marqué du fragment *Ase*I de 110 kb, élué à partir d'un profil PFGE de NSA27. Les inserts de tous les cosmides ainsi isolés correspondaient à des fragments *Bam*HI déjà cartographiés dans la souche sauvage et provenant du fragment *Ase*I J. La carte de restriction *Bam*HI du fragment *Ase*I de 110 kb a alors été dressée à l'aide du chevauchement de ces cosmides (Figure R-4).



**Figure R-4 : Cartographie *Bam*HI du fragment *Ase*I de 110 kb de la souche NSA27.**

Le symbolisme utilisé est identique à celui de la Figure R-2. À gauche profil de restriction des cosmides 49.2, 56.2 et 24.1. Le v indique la bande à 6,8 kb correspondant au vecteur. À droite, la localisation des trois cosmides portant le fragment de fusion de 11 kb est indiquée sous la cartographie *Bam*HI du fragment *Ase*I de 110 kb. La séquence dupliquée en orientation inverse est symbolisée par des flèches. La taille du fragment de 110 kb est compatible avec la fusion inverse de la séquence non délétée de 70 kb du fragment *Ase*I J sauvage de 95 kb et d'une partie de 40 kb provenant du même locus chromosomique.

Ainsi, le fragment de 110 kb peut être interprété comme résultant de la fusion de deux fragments *Ase*I-J partiellement délétés : l'un délété jusque dans le fragment *Bam*HI de 16 kb et mesurant 70 kb, l'autre délété jusque dans le fragment *Bam*HI de 15 kb et mesurant 40 kb. Le nouveau fragment de 11 kb (cloné dans le cosmide 49.2) consiste donc en la fusion de deux fragments *Bam*HI de 16 et 15 kb partiellement délétés.

La structure chromosomique du mutant NSA27 a également été analysée à l'aide de l'enzyme de restriction *Dra*I (la seconde enzyme utilisée pour la cartographie du chromosome de *S. ambofaciens*, (Leblond *et al.*, 1996)).

Ainsi, le chromosome du mutant NSA27 présente une délétion des deux fragments terminaux *Dra*I IV-730 kb et I-2100 kb (Figure R-5, le fragment *Dra*I IV n'est pas visible à l'aide de cette résolution). De plus, un nouveau fragment *Dra*I d'environ 3 Mb a été détecté dans le profil mutant (Figure R-5). Selon la délétion caractérisée à l'aide de la cartographie *Ase*I, 1500 kb d'ADN du fragment *Dra*I I-2100 kb doivent être conservés. Ainsi, le nouveau fragment *Dra*I de 3 Mb a été interprété comme la fusion en orientation inverse de deux fragments *Dra*I I-2100 kb partiellement délétés. L'hybridation du cosmide 9D10, spécifique du fragment *Dra*I I-2100 kb, permet en effet de révéler ce nouveau fragment de 3 Mb (Figure R-5).



A-I.1 ♂ Structure chromosomique fusionnée des mutants *S. ambofaciens* NSA27 et NSA65

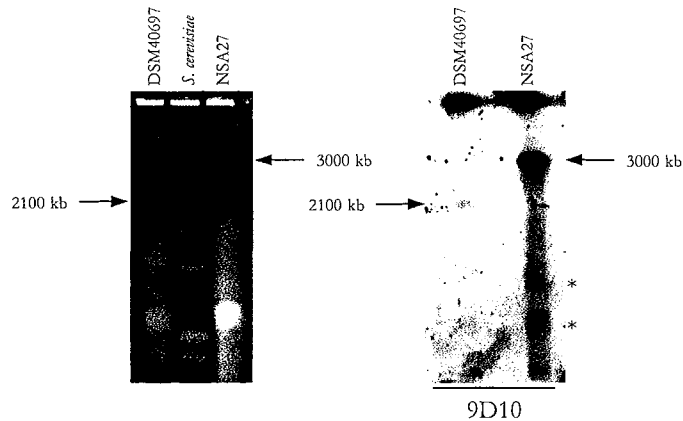
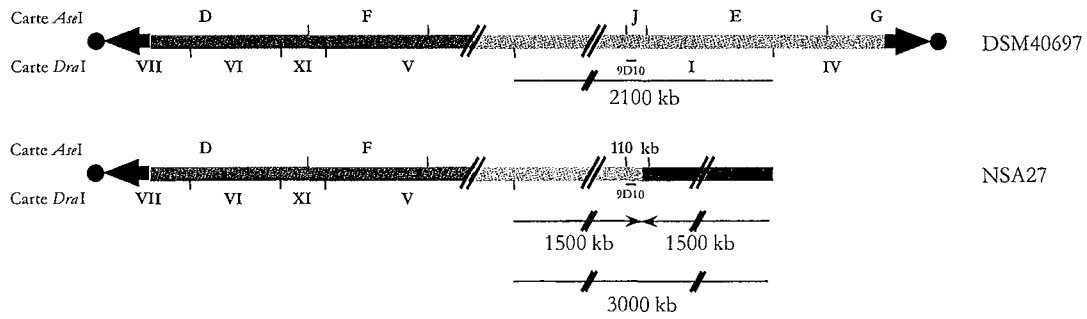


Figure R-5 : Comparaison des cartes de restriction *DrrI* sauvage et mutante NSA27.

Le symbolisme utilisé est le même que celui de la Figure R-2. Seules les tailles des bandes d'intérêt sont indiquées. Le marqueur de tailles utilisé correspond aux chromosomes de la levure *Saccharomyces cerevisiae*. L'hybridation a été réalisée à l'aide de l'ADN marqué du cosmide 9D10 sur les profils de restriction montrés à gauche. Les astérisques indiquent des signaux d'hybridation sur des bandes non visibles sur le profil de restriction. Elles pourraient correspondre à une hybridation sur de l'ADN dégradé. Les conditions de migration en PFGE sont les suivantes : 3 V.cm<sup>-1</sup>, 90 h en gel d'agarose à 0,6% avec des temps de pulsation variant de 200 s à 500 s, angle inclus 120°.

L'ensemble de ces données suggérerait alors que la structure chromosomique du mutant NSA27 résultait de la fusion en orientation inversée de deux chromosomes partiellement délétés sur le même bras chromosomique.

### A-I.1.2 - LE MUTANT *S. AMBOFACIENS* NSA65

Une situation similaire a été caractérisée dans la souche NSA65. Une délétion du bras droit d'environ 1330 kb, affecte les fragments *AseI*-G et *AseI*-E mais aucun nouveau fragment *AseI* n'a pu être détecté (Figure R-6).

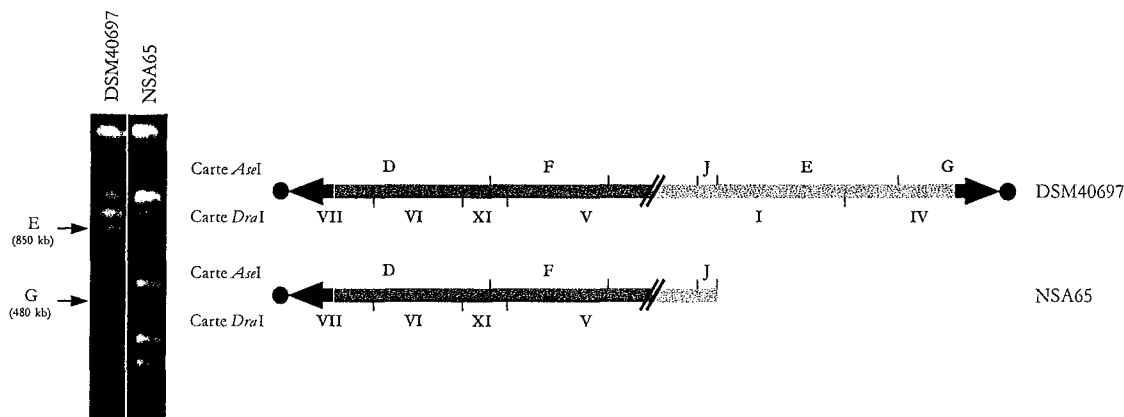


Figure R-6 : Structure chromosomique de la souche NSA65.  
Le symbolisme utilisé est le même que dans la Figure R-2.

La borne de délétion a été localisée dans le fragment *Bam*HI de 12 kb porté par le cosmide 2B7 et un nouveau fragment de 18,5 kb a pu être détecté (Figure R-7). Suite à des analyses de restriction et des hybridations avec des sondes spécifiques, ce fragment de jonction a été interprété comme la fusion de deux fragments *Bam*HI de 12 kb partiellement délétés (résultats non montrés). Ainsi, cette structure générerait un fragment de fusion *Ase*I de 94,5 kb, qui ne serait pas résolu du fragment *Ase*I J-95 kb sauvage dans les conditions de PFGE utilisées.

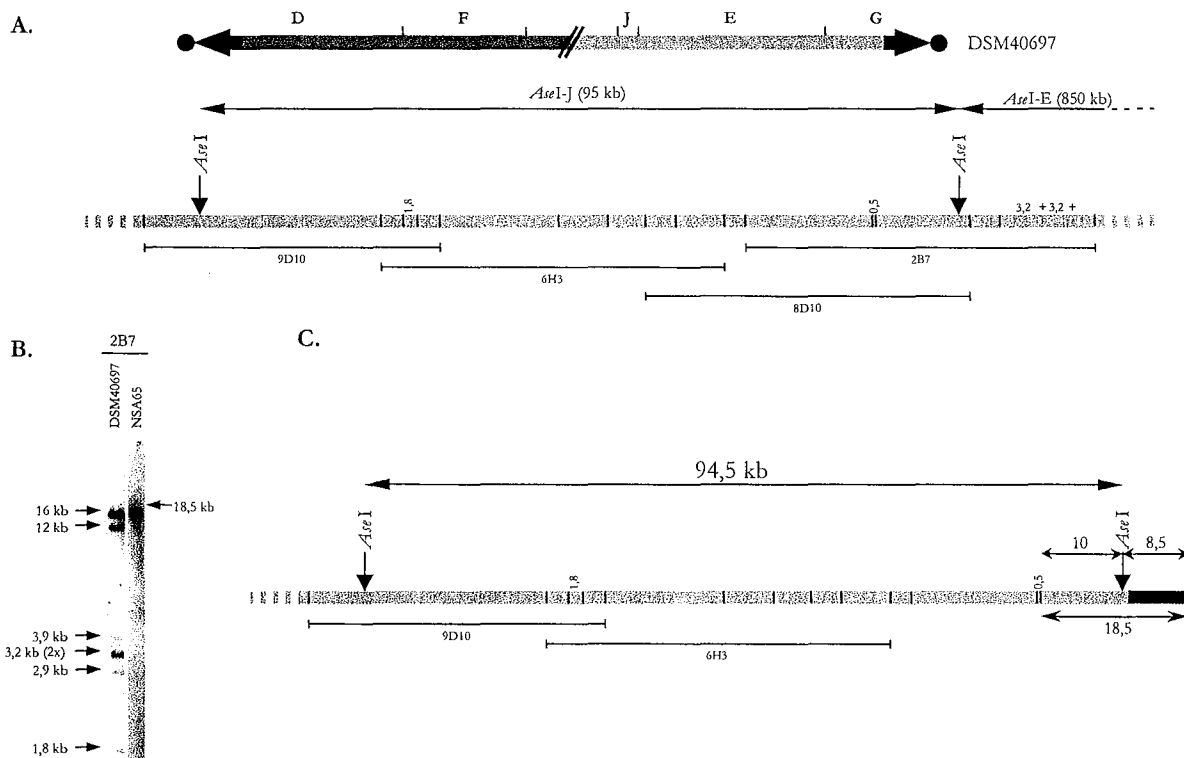


Figure R-7 : Mise en évidence du réarrangement dans la souche *S. ambifaciens* NSA65.

A. Cartographies *Ase*I du chromosome et *Bam*HI du fragment *Ase*I-J de *S. ambifaciens* DSM40697.

B. Hybridation de l'ADN marqué du cosmide 2B7 un profil de restriction *Bam*HI des souches sauvage et mutante NSA65.

C. Cartographie du réarrangement mis en évidence chez *S. ambifaciens* NSA65.

Enfin, comme pour NSA27, le profil de restriction *Dra*I de la souche NSA65 révèle un nouveau fragment d'environ 3 Mb, compatible avec la duplication de la partie intacte conservée du fragment *Dra*I I-2100 kb (Figure R-8).

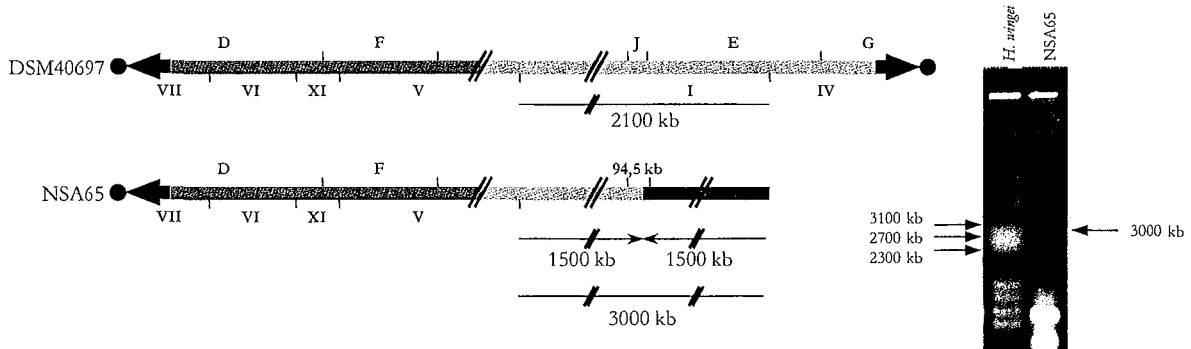


Figure R-8 : Comparaison des cartes de restriction *Dra*I sauvage et mutante NSA65.

Le symbolisme utilisé est le même que celui de la Figure R-2. Seules les tailles des bandes d'intérêt sont indiquées. Le marqueur de tailles utilisé correspond aux chromosomes de la levure *Hansenula wingei*. Les conditions de migration en PFGE sont les suivantes : 3 V.cm<sup>-1</sup>, 48 h en gel d'agarose 0,6%, des temps de pulsation variants de 200 s à 500 S et un angle inclus de 94°.

L'ensemble de ces résultats suggère donc que la structure chromosomique des mutants NSA65 et NSA27 est issue d'un événement de fusion entre deux chromosomes partiellement délétés. Les chromosomes mutants auraient ainsi une taille d'environ 13 Mb, comparativement au chromosome de 8 Mb chez la souche sauvage, avec des régions dupliquées d'environ 6500 kb constituant de nouvelles TIR (Figure R-9).

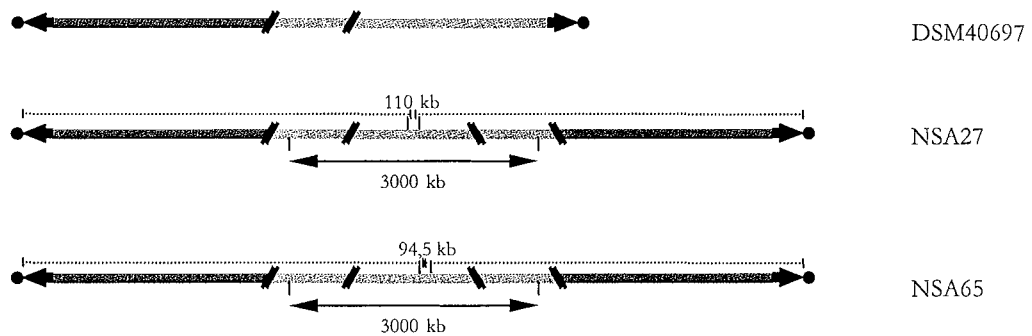
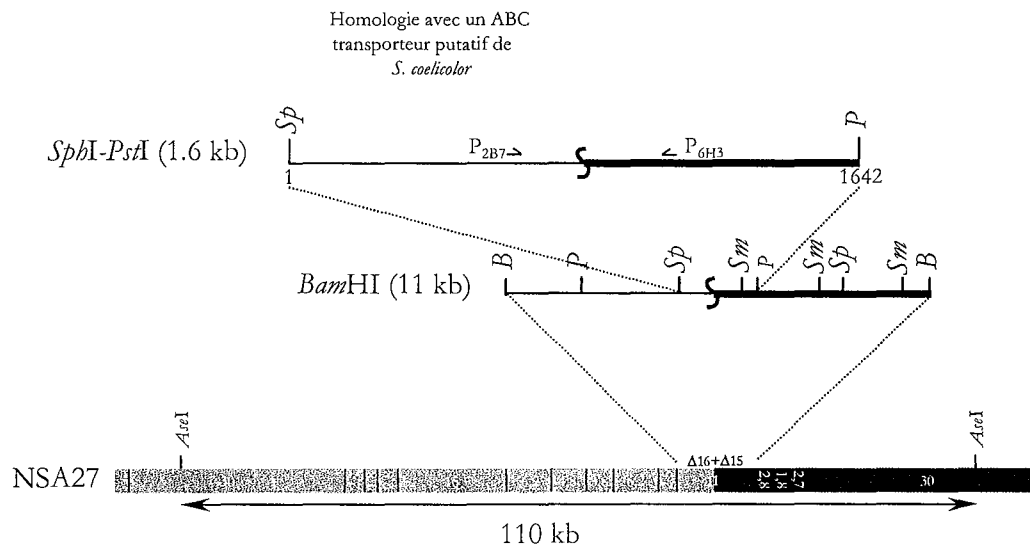


Figure R-9 : Structures chromosomiques des mutants *S. ambofaciens* NSA27 et NSA65.

Seuls les régions contenant le réarrangement sont indiquées. L'étendue de la duplication est indiquée par les traits pointillés. Le symbolisme utilisé est le même que celui de la Figure R-2.

## A-I.2 $\&$ ANALYSE DES SEQUENCES IMPLIQUEES DANS LE REARRANGEMENT CHEZ *S. AMBOFACIENS* NSA27

Après avoir réalisé une cartographie fine des bornes du réarrangement, le point de fusion du chromosome de *S. ambofaciens* NSA27 a été caractérisé par séquençage d'un fragment *Pst*I-*Sph*I de 1,6 kb, cartographié à l'intérieur du fragment *Bam*HI de 11 kb portant la jonction (Figure R-10). Une forte homologie, de 95% d'identité sur les 856 premiers nucléotides, a été détectée avec une partie d'une ORF de *S. coelicolor* A3(2) codant un ABC transporteur putatif. Cette homologie est alors brusquement interrompue, le reste de la séquence ne présentant aucune identité notable avec les banques de données, suggérant la proximité du point de recombinaison.



**Figure R-10 : Analyse du point de recombinaison chez *S. ambofaciens* NSA27.**

Carte de restriction du fragment *Bam*HI de 11 kb incluant le point de recombinaison (B: *Bam*HI; P: *Pst*I, Sp: *Sph*I; Sm: *Sma*I). Ce fragment consiste en la fusion de deux fragments *Bam*HI partiellement délétés de 16 kb et 15 kb, provenant respectivement des cosmides 2B7 (trait fin) et 6H7 (trait épais). La zone d'homologie détectée dans la séquence du fragment *Sph*I-*Pst*I (1642 nucléotides, numéro d'accèsion GenBank AY227908) est indiquée, ainsi que la position des oligonucléotides (*P*<sub>6H3</sub> and *P*<sub>2B7</sub>) utilisés pour le séquençage des bornes sauvages impliquées dans le réarrangement (numéros d'accèsion GenBank AY227910 pour 6H3 et AY227909 pour 2B7).

Deux oligonucléotides (*P*<sub>2B7</sub> and *P*<sub>6H3</sub>, Figure R-10) ont alors été dessinés de part et d'autre de ce point de recombinaison putatif, et les séquences sauvages impliquées dans l'événement de recombinaison ont été déterminées à partir des cosmides 2B7 et 6H3. La séquence déduite du cosmide 2B7 (26 nt) confirme l'homologie avec l'ABC transporteur précédemment identifié alors qu'aucune homologie significative n'a pu être déduite de l'analyse de la séquence obtenue à partir du cosmide 6H3 (503 nt).

L'alignement de la séquence de jonction avec les séquences sauvages a alors montré que le réarrangement caractérisé dans la souche NSA27 est issu d'un événement de recombinaison illégitime entre des microhomologies de 6 pb (Figure R-11).

```
WT (6H3)  GTACGGAGGTGGCCCGTCTGGAGGCCCAGGACAACT
NSA27    GCGCGGCTTGGGCCCGTCTGGAGGCCCAGGACAACT
WT (2B7) GCGCGGCTTGGGCCCGTCTGGTAGCGCGAGCAGCCG
```

**Figure R-11 : Alignement des séquences sauvages impliquées dans le réarrangement et de la séquence au point de recombinaison chez NSA27.**

La séquence de 6 pb impliquée dans l'événement de recombinaison illégitime est encadrée. Les caractères en traits fins et en traits gras correspondent aux nucléotides provenant de la séquence des cosmides 2B7 et 6H3 respectivement.

- CHAPITRE A-II -

LA FUSION CHROMOSOMIQUE : UN  
ETAT MUTATEUR ?

## A-II.1 $\sphericalangle$ FUSION DE CHROMOSOMES ET INSTABILITE CHROMOSOMIQUE

### A-II.1.1 – INSTABILITE CHROMOSOMIQUE DES MUTANTS *S. AMBOFACIENS* NSA27 ET NSA65

Afin d'analyser la stabilité génomique des mutants présentant une fusion de deux chromosomes partiellement délétés, la structure chromosomique de descendants des souches mutantes *S. ambofaciens* NSA27 et NSA65, a été caractérisée par plusieurs approches :

- La comparaison des profils de restriction *AseI* a permis de caractériser des grands réarrangements additionnels,
- L'hybridation de l'ADN marqué du cosmide AD91 (localisé dans les TIR du chromosome) sur les profils de restriction *AseI* a permis de caractériser les réarrangements des extrémités,
- Le traitement ou non de l'ADN à la protéinase lors des expériences de PFGE révéla le caractère linéaire ou non du chromosome analysé,
- Les hybridations de l'ADN marqué des cosmides 49.2, 6H3 ou 9D10 sur les profils de restriction *AseI* révélèrent les réarrangements des séquences portant la fusion.

L'ensemble des structures chromosomiques déduites de cette analyse est résumé dans les Tableaux R-1 et R-2. Ces différentes expériences ont permis de montrer qu'un réarrangement de la structure chromosomique fusionnée parentale était très fréquent dans les descendances des deux mutants NSA27 et NSA65. En effet, 11/19 et 7/15 des colonies testées pour les souches NSA27 et NSA65, respectivement, présentent une structure non parentale (NP), pouvant correspondre à un chromosome, soit linéaire (L), soit circulaire (C).

Une structure chromosomique circulaire a ainsi pu être caractérisée chez 5 descendants de NSA27 et 3 de NSA65. En ce qui concerne les chromosomes linéaires, deux types de structures ont été distingués :

- Des chromosomes linéaires présentant une fusion chromosomique différente de la fusion parentale (NF pour Nouvelle Fusion). Cette structure a été observée chez 1 et 3 descendants de NSA27 et NSA65 respectivement.
- Des chromosomes linéaires ayant perdu la structure fusionnée (PF pour Perte de la Fusion). Ces chromosomes, retrouvés chez 5 descendants de NSA27 et 1 descendant de NSA65, présentent alors une grande variabilité de la taille des TIR, comprise entre 105 et 1300 kb.

Souche	Type	Structure chromosomique vraisemblable	Taille du chromosome (kb)	Taille de la TIR (kb)
Sauvage			8 000	210
NSA27			13 200	6 500
NSA27 <sub>S3</sub>	P		13 200	6 500
NSA27 <sub>S6</sub>			13 200	6 500
NSA27 <sub>S9</sub>			13 200	6 500
NSA27 <sub>S10</sub>			13 200	6 500
NSA27 <sub>S12</sub>			13 200	6 500
NSA27 <sub>S15</sub>			13 200	6 500
NSA27 <sub>S16</sub>			13 200	6 500
NSA27 <sub>S19</sub>			13 200	6 500
NSA27 <sub>S1</sub>		?	12 000	0
NSA27 <sub>S2</sub>	C		12 000	0
NSA27 <sub>S4</sub>			12 000	0
NSA27 <sub>S11</sub>			12 000	0
NSA27 <sub>S14</sub>			6 500	0
NSA27 <sub>S17</sub>	NF		13 200	6 500
NSA27 <sub>S5</sub>	PF		7 985	1 300
NSA27 <sub>S20</sub>			7 985	1 300
NSA27 <sub>S8</sub>			7 730	1 140
NSA27 <sub>S13</sub>			7 730	1 140
NSA27 <sub>S18</sub>			7 055	370

Tableau R-1 : Récapitulatif des structures chromosomiques des descendants de la souche *S. ambofaciens* NSA27.

Les divers types de structures (P pour Parentale, C pour Circulaire, NF pour Nouvelle Fusion et PF pour Perte de la Fusion), la taille des chromosomes ainsi que des répétitions terminales inversées sont indiquées. Un point d'interrogation signifie que la structure chromosomique reste à préciser.

Souche	Type	Structure chromosomique vraisemblable	Taille du chromosome (kb)	Taille de la TIR (kb)
Sauvage			8 000	210
NSA65			13 340	6 500
NSA65 <sub>S2</sub>	P		13 340	6 500
NSA65 <sub>S3</sub>			13 340	6 500
NSA65 <sub>S5</sub>			13 340	6 500
NSA65 <sub>S6</sub>			13 340	6 500
NSA65 <sub>S11</sub>			13 340	6 500
NSA65 <sub>S14</sub>			13 340	6 500
NSA65 <sub>S15</sub>			13 340	6 500
NSA65 <sub>S16</sub>			13 340	6 500
NSA65 <sub>S1</sub>	C	?	?	0
NSA65 <sub>S8</sub>			6 100	0
NSA65 <sub>S18</sub>		?	?	0
NSA65 <sub>S4</sub>	NF		13 300	6 600
NSA65 <sub>S13</sub>			13 300	6 600
NSA65 <sub>S20</sub>			13 300	6 600
NSA65 <sub>S19</sub>	PF		6 770	105

Tableau R-2 : Récapitulatif des structures chromosomiques des descendants de la souche *S. ambofaciens* NSA65.

Le symbolisme est identique à celui du Tableau R-1.



---

**Ainsi, la structure fusionnée semble être très instable. Les mutants NSA27 et NSA65 peuvent être considérés comme des souches mutatrices, générant une grande variabilité de structures chromosomiques dans leur descendance et pourraient ainsi refléter les mécanismes d'évolution des génomes.**

### **A-II.1.2 - UN CYCLE D'INSTABILITE CHROMOSOMIQUE**

L'héritabilité des différentes structures chromosomiques précédemment décrites (P pour parentale, NF pour Nouvelle Fusion et PF pour Perte de la Fusion) a ensuite été analysée dans la descendance de 4 dérivés des mutants NSA27 (NSA27<sub>S10</sub>, NSA27<sub>S17</sub>, NSA27<sub>S5</sub> et NSA27<sub>S18</sub>) et NSA65 (NSA65<sub>S4</sub>, NSA65<sub>S5</sub>, NSA65<sub>S20</sub> et NSA65<sub>S19.S3</sub>).

Les différentes conclusions émises à l'issue de cette analyse sont résumées dans le Tableau R-3. Cette analyse a alors permis d'associer les structures chromosomiques à une instabilité chromosomique particulière.

En effet, l'état fusionné, retrouvé dans les deux types de structures P et NF, paraît strictement corrélé à une forte instabilité génomique puisque 10 à 67% des descendants analysés présentent un réarrangement additionnel par rapport à la structure chromosomique du mutant dont ils dérivent.

Au contraire, les descendants ayant perdu la fusion chromosomique ancestrale présentent un profil d'héritabilité de leur structure chromosomique indistinguable de celui de la souche sauvage puisque aucun réarrangement n'a été observé chez les 10 descendants des 5 souches analysées.

		Instabilité génomique				
Souche	Taille des TIR (kb)	n	P	NP		
				C	L	
					PF	NF
DSM40697	210	24	24	0	0	0
NSA27	6500	19	8	5	5	1
NSA65	6500	15	8	3	1	3
NSA27 <sub>S10</sub>	6500	10	9	1	0	0
NSA65 <sub>S5</sub>	6500	10	9	1	0	0
NSA27 <sub>S17</sub>	6500	20	9	4	7	0
NSA65 <sub>S4</sub>	6500	9	3	4	2	0
NSA65 <sub>S20</sub>	6500	10	8	2	0	0
NSA27 <sub>S5</sub>	1300	10	10	0	0	0
NSA27 <sub>S18</sub>	370	10	10	0	0	0
NSA65 <sub>S19S3</sub>	1005	10	10	0	0	0

Tableau R-3 : Caractérisation génomique des souches *S. ambofaciens* sauvage et mutantes NSA27 et NSA65 ainsi que de leurs descendants.

Les différentes souches sont classées en fonction de leur structure chromosomique : P pour Parentale, NF pour Nouvelle Fusion (i.e. conservation de la structure fusionnée mais avec un nouveau fragment de jonction *AseI*) et PF pour Perte de la Fusion (i.e. conservation de la structure linéaire du chromosome mais perte de la structure fusionnée parentale). La taille des TIR est donnée en kilobases et est déduite des expériences de PFGE et d'hybridations. NP est utilisé pour signifier une structure Non Parentale, C pour Circulaire. Les chromosomes linéaires (L) peuvent être des deux types PF ou NF.

Ainsi, l'instabilité chromosomique, détectée chez les mutants NSA27 et NSA65, est héritable si la duplication parentale est maintenue. La perte de la structure fusionnée (perte de la majeure partie de la duplication) est corrélée à la "stabilisation" relative de la structure chromosomique.

## A-II.2 $\otimes$ FUSION DE CHROMOSOMES ET INSTABILITE PHENOTYPIQUE

### A-II.2.1 – INSTABILITE PHENOTYPIQUE DES MUTANTS *S. AMBOFACIENS* NSA27 ET NSA65 ET DE LEURS DESCENDANTS

L'instabilité phénotypique des mutants NSA27 et NSA65 ainsi que de leurs descendants a été alors estimée, après 7 jours de croissance, en comptant les variants phénotypiques (morphologique et pigmentation) parmi des colonies isolées issues de chaque sporée (Tableau R-4). Dans nos expériences, l'instabilité phénotypique de la souche sauvage a pu être estimée à 1,4%, ce qui est concordant avec les précédentes études (Leblond *et al.*, 1991; Fischer *et al.*, 1997). L'instabilité phénotypique des mutants s'élève respectivement à 36,4% et 30,2% pour NSA65 et NSA27 (Tableau R-4). Ainsi, l'instabilité phénotypique des mutants présentant une fusion chromosomique est bien supérieure à celle de la souche sauvage.

De plus, une corrélation stricte a pu être observée entre le statut chromosomique et le niveau d'instabilité phénotypique dans la descendance des deux souches mutantes. En effet, les souches de type parental (P) ou portant une nouvelle fusion (NF) présentent des niveaux d'instabilité phénotypique élevés (de 22,5% à 30,8%) alors que celles ayant perdu la structure fusionnée (PF) retrouvent une instabilité phénotypique de type sauvage (comprise entre 0,18 et 2,1%).

Souche	Taille des TIR (kb)	Instabilité génomique					Instabilité phénotypique
		n	P	NP			
				C	L		
					PF	NF	
DSM40697	210	24	24	0	0	0	1.4 $\pm$ 1.1 %
NSA27	6500	19	8	5	5	1	30.2 $\pm$ 8.1 %
NSA65	6500	15	8	3	1	3	36.4 $\pm$ 8.1 %
NSA27 <sub>S10</sub>	6500	10	9	1	0	0	27 $\pm$ 6.3%
NSA65 <sub>S5</sub>	6500	10	9	1	0	0	26 $\pm$ 6.5 %
NSA27 <sub>S17</sub>	6500	20	9	4	7	0	22.5 $\pm$ 5.9 %
NSA65 <sub>S4</sub>	6500	9	3	4	2	0	30.8 $\pm$ 5.8 %
NSA65 <sub>S20</sub>	6500	10	8	2	0	0	26.9 $\pm$ 7.0 %
NSA27 <sub>S5</sub>	1300	10	10	0	0	0	0.45 $\pm$ 0.9 %
NSA27 <sub>S18</sub>	370	10	10	0	0	0	0.7 $\pm$ 1.0 %
NSA65 <sub>S1983</sub>	1005	10	10	0	0	0	0.18 $\pm$ 0.4 %

Tableau R-4 : Caractérisation génomique et phénotypique des souches *S. ambofaciens* sauvage et mutantes NSA27 et NSA65 ainsi que de leurs descendants.

Le symbolisme utilisé est identique à celui du Tableau R-3. Les pourcentages d'instabilité phénotypique correspondent aux fréquences de colonies présentant un phénotype non parental. Le nombre de colonies analysées est compris entre 140 et 487 selon la souche considérée. Les souches apparaissant en vert présentent des caractéristiques spécifiques à l'état fusionné, alors que la couleur rouge symbolise des comportements de type sauvage.

L'ensemble de ces données permet de classer les différentes souches analysées en deux catégories : la première est composée de souches présentant des répétitions terminales inversées de taille inférieure à 1300 kb. Elles présentent une instabilité chromosomique et phénotypique de type sauvage (en couleur rouge dans le Tableau R-4). La seconde catégorie est composée des souches présentant une duplication d'environ 6500 kb. Chez ces souches, l'instabilité chromosomique est corrélée à une forte instabilité phénotypique (en couleur verte dans le Tableau R-4).

## A-II.2.2 – CORRELATION ENTRE ETAT FUSIONNE ET DEFAUT DE SPORULATION

### A-II.2.2.1 – ANALYSE DU CONTENU EN ADN DES SPORES PAR COLORATION AU DAPI

La répartition de la quantité d'ADN dans les spores des souches sauvage, NSA27, NSA65, NSA27<sub>S17</sub>, NSA27<sub>S5</sub>, NSA27<sub>S18</sub> et NSA65<sub>S19S3</sub> a été étudiée par mesure de l'intensité de fluorescence suite à une coloration de l'ADN au DAPI (Kieser *et al.*, 2000).

Des chaînettes dont les spores sont situées dans le même plan focal ont été sélectionnées. Les répartitions des différentes intensités ainsi mesurées sont reportées sur la Figure R-12.

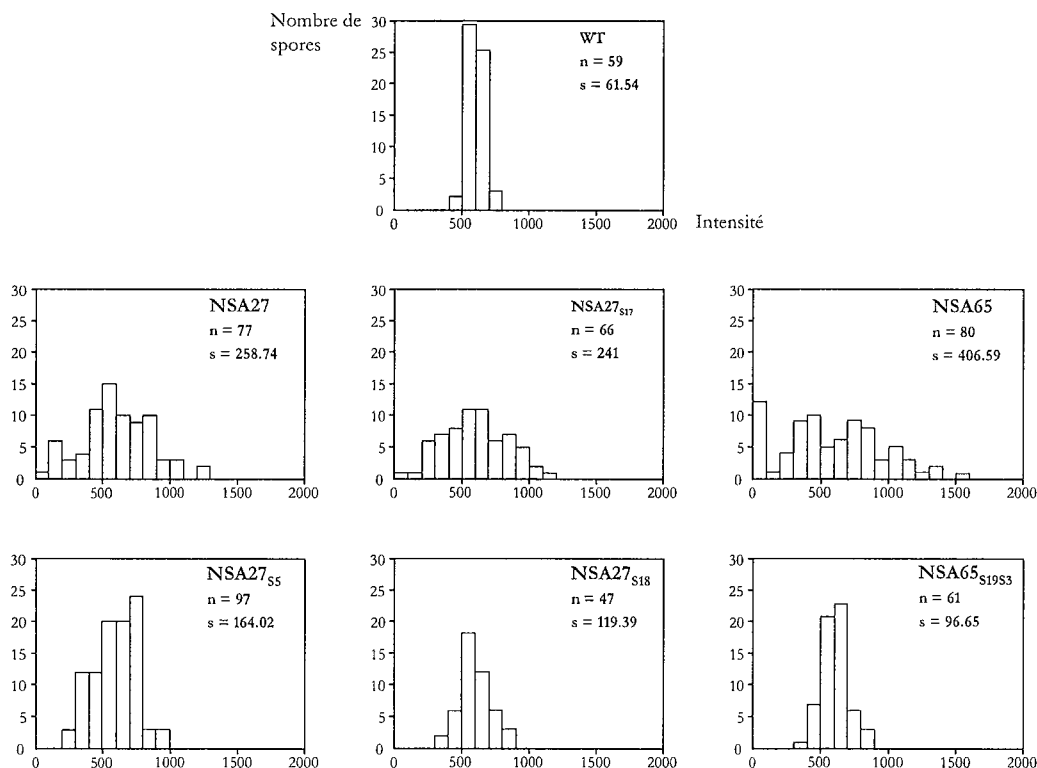


Figure R-12 : Distribution de la quantité d'ADN dans les spores des souches sauvage (en haut) et mutantes portant une fusion chromosomique (au milieu) et de descendants ayant perdu la fusion parentale (en bas).

La déviation standard est donnée par la valeur  $\sigma$ . Toutes les distributions ont été centrées par rapport à une valeur d'intensité arbitraire (exprimée en pixels par spore) de 600.

Toutes les distributions, à l'exception de celle de la souche NSA65, sont de type normal ( $P < 0,01$ ) permettant alors la comparaison de leur déviation standard ( $\sigma$ ). Ainsi, il apparaît que les distributions des souches NSA27 and NSA27<sub>S17</sub> sont plus larges que celles de la souche sauvage et des autres descendants analysés, révélant une forte hétérogénéité du contenu en ADN des spores chez ces deux souches.

Le fait que la distribution de la souche NSA65 ne soit pas de type Gaussien ( $P < 0,01$ ) révèle l'existence de types distincts de spores (par exemple, des spores dépourvues en ADN) mais ne permet pas les comparaisons avec les autres distributions.

Enfin, les descendants des souches mutantes NSA27 et NSA65 ayant perdu la fusion chromosomique parentale présentent une distribution de l'intensité de fluorescence des spores plus proche de celle de la souche sauvage que de celle de leur souche parentale ( $P < 0,01$ ).

L'ensemble de ces données confirme les observations directes des clichés photographiques, révélant une grande hétérogénéité de la distribution du contenu en ADN des spores chez les mutants portant une fusion chromosomique mais aussi le retour à une répartition homogène associé à la perte de la fusion (Figure R-13).

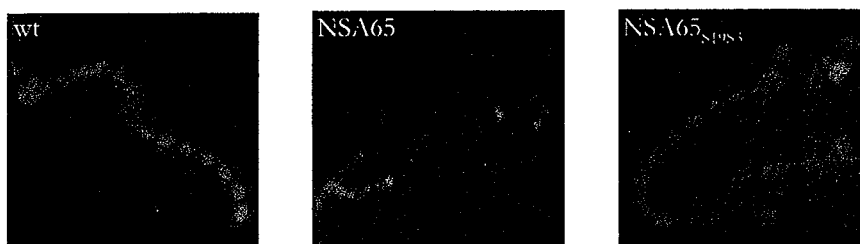


Figure R-13 : Etude par coloration au DAPI de la quantité d'ADN dans les spores des souches sauvage (à gauche) et mutante portant une fusion chromosomique (NSA65, au milieu) et d'un descendant ayant perdu la fusion parentale (à droite, NSA65<sub>S19S3</sub>).

Photographies en lumière transmise et en fluorescence.

Les grossissements sont de 1000 fois.

#### A-II.2.2.2 – ANALYSE DE L'ULTRASTRUCTURE DES SPORES PAR MICROSCOPIE ELECTRONIQUE

L'observation par microscopie électronique à transmission de chaînettes de spores révèle la présence de spores « vides », ou spores dépourvues de contenu, ainsi qu'une grande hétérogénéité dans la taille des spores chez les souches mutantes NSA27 et NSA65 (Figure R-14.B). De plus, des Figures aberrantes de sporulation n'ont été observées que chez ces deux souches (Figure R-14.D). Un composant cellulaire, pouvant correspondre à de l'ADN, est en effet fréquemment observé dans la zone de septation et semblerait empêcher une correcte séparation des spores adjacentes (Figure R-14.D).

Au contraire, chez la souche sauvage ou chez les descendants ayant perdu la fusion chromosomique, ni de Figure aberrante de sporulation ni spore « vide » ne sont observées (exemple de la souche NSA27<sub>S19S3</sub> en Figure R-14 C.).

Ainsi, le défaut de sporulation observé est strictement associé ( $P < 0,05$ ) à la fusion chromosomique, puisqu'il disparaît de façon concomitante à la perte de cette structure fusionnée. Enfin, la taille des TIR (variant de 210 kb chez la souche sauvage à 1005 kb chez la souche mutante NSA65<sub>S19S3</sub>) ne semble pas influencer la capacité de sporulation.

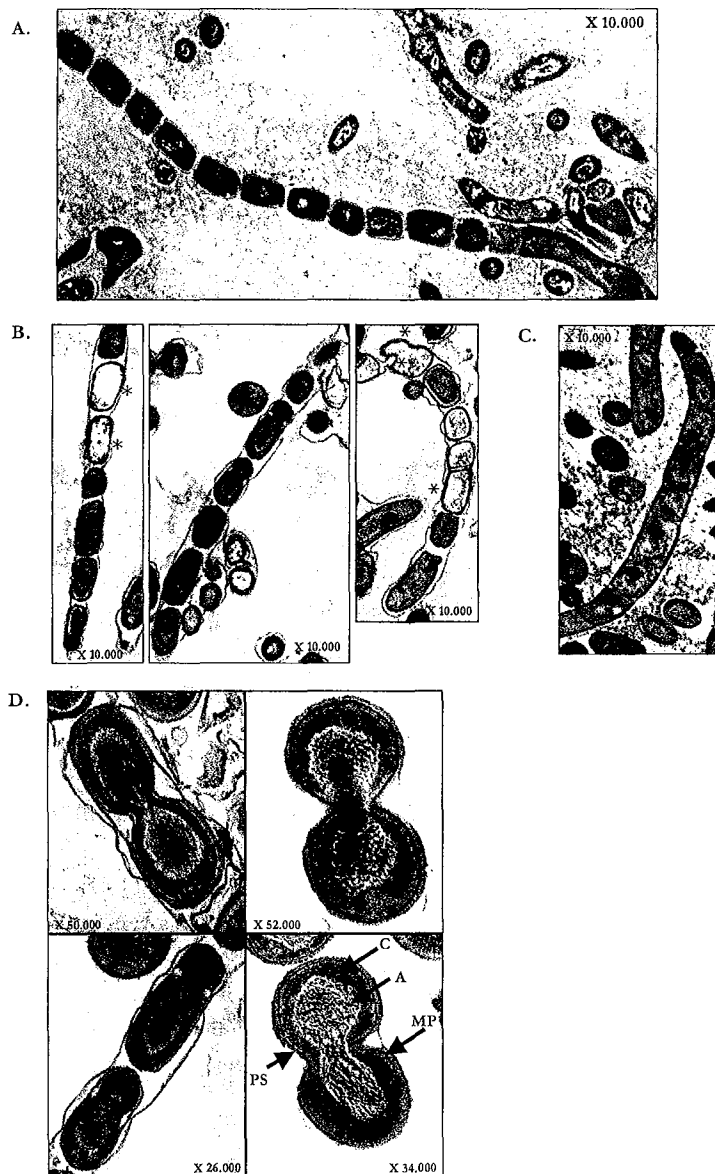


Figure R-14 : Etude en microscopie électronique à transmission de la morphologie des spores des souches *S. ambifaciens* DSM40697 (A.), NSA65 (B.) et NSA65<sub>S19S3</sub> (C.).

Un astérisque indique une spore « vide ». Des exemples de Figures aberrantes de sporulation sont présentés en D. (NSA65 à gauche et NSA27 à droite). Le cytoplasme (C), la paroi de la spore (PS), la membrane plasmique (MP), ainsi que l'ADN (A) sont indiqués. Les grossissements sont indiqués sur chaque photographie.

A-II.2.2.3 – ANALYSE DU TAUX DE GERMINATION

Le défaut de sporulation a enfin été confirmé par la détermination du taux de germination (Tableau R-5). L'analyse des ces taux de germination chez les diverses souches analysées a permis de dégager une tendance : alors que les souches portant une fusion chromosomique présentent de faibles taux de germination, les souches l'ayant perdu retrouvent des taux de type sauvage.

Souche	Taille des TIR (kb)	Instabilité génomique					Instabilité phénotypique	Taux de germination	
		n	P	NP					
				C	L				
					PF	NF			
DSM40697	210	24	24	0	0	0	1.4 $\pm$ 1.1 %	66%	
NSA27	6500	19	8	5	5	1	30.2 $\pm$ 8.1 %	9%	
NSA65	6500	15	8	3	1	3	36.4 $\pm$ 8.1 %	25.6%	
NSA27 <sub>s10</sub>	P	6500	10	9	1	0	0	27 $\pm$ 6.3%	6%
NSA65 <sub>s5</sub>		6500	10	9	1	0	0	26 $\pm$ 6.5 %	16%
NSA27 <sub>s17</sub>	NF	6500	20	9	4	7	0	22.5 $\pm$ 5.9 %	5%
NSA65 <sub>s4</sub>		6500	9	3	4	2	0	30.8 $\pm$ 5.8 %	8%
NSA65 <sub>s20</sub>		6500	10	8	2	0	0	26.9 $\pm$ 7.0 %	16%
NSA27 <sub>s5</sub>	PF	1300	10	10	0	0	0	0.45 $\pm$ 0.9 %	77%
NSA27 <sub>s18</sub>		370	10	10	0	0	0	0.7 $\pm$ 1.0 %	71%
NSA65 <sub>s19S3</sub>		1005	10	10	0	0	0	0.18 $\pm$ 0.4 %	76%

Tableau R-5 : Caractérisation génomique et phénotypique des souches *S. ambofaciens* sauvage et mutantes NSA27 et NSA65 ainsi que de leurs descendants.

Le symbolisme utilisé est identique à celui du Tableau R-4. Les taux de germination sont donnés après 4 jours de croissance.

B - SECONDE PARTIE :

UN EVENEMENT DE DUPLICATION  
GENIQUE DANS LES REGIONS  
TERMINALES DU CHROMOSOME DE  
*STREPTOMYCES AMBOFACIENS*



S.C.D. - Université de Nancy  
BIBLIOTHÈQUE DES SCIENCES  
Rue du Jardin Botanique  
54600 VILLERS-LES-NANCY

- CHAPITRE B-I -

CARACTERISATION DES GENES *HAS*

En plus de révéler la dynamique de la plasticité du chromosome de *S. ambofaciens*, la caractérisation de la structure chromosomique des souches mutantes NSAH135 et NSA855 (Fischer *et al.*, 1998b) a également permis d'étudier un événement de duplication fixé au cours de l'évolution. En effet, lors de l'identification des séquences impliquées dans les événements de remplacement de bras chromosomiques présents chez les mutants NSAH135 et NSA855, une forte homologie entre deux loci du chromosome de la souche sauvage *S. ambofaciens* DSM40697 localisés à 480 et 850 kb des extrémités a été décelée. La recherche de phase ouverte de lecture dans ces régions a alors révélé la présence d'une duplication de gènes, nommés *hasR* et *hasL*, codant des facteurs  $\sigma$  alternatifs.

*Les résultats de cette seconde partie sont exposés dans la publication en préparation pour Journal of Bacteriology (Roth et al., 2003) (page 356).*

## **B-I.1 ☞ LES GENES *HASR* ET *HASL* CHEZ LA SOUCHE SAUVAGE *STREPTOMYCES AMBOFACIENS* DSM40697**

### **B-I.1.1 – LES FACTEURS $\sigma^{HAS}$**

La région homologue impliquée dans le réarrangement caractérisé chez les mutants NSAH135 et NSA855 débute 4 nucléotides en amont du codon d'initiation de la traduction des gènes *has* et s'arrête 8 nucléotides en amont du codon d'arrêt de la traduction. Ainsi, les gènes *hasL* et *hasR* présentent 98% d'identité au niveau nucléotidique (14 différences sur 846 pb, Figure R-15) et 97% d'identité protéique (Figure R-16). Les deux protéines putatives de 281 acides aminés ne diffèrent en effet que par 8 positions dont 3 constituent leur extrémité C-terminale.

<i>hasR</i>	1	ATGGCGaCGATGACCGCACGGATCACCGAGCACACCACCACGGGCATGCCGGAGGTGGCG	60
<i>hasL</i>	1	ATGGCGACGATGACCGCA GGATCACCGAGCACACCACCACGGGCATGCCGGAGGTGGCG	60
<i>hasR</i>	61	GACCCGTCCAAGGTGGCGCCCAAGGACGCGCGGGAGTTGTCTGAAGCTCTTCTTCCAGCAG	120
<i>hasL</i>	61	GACCCGTCCAAGGTGGCGCCCAAGGACGCGCGGGAGTTGTCTGAAGCTCTTCTTCCAGCAG	120
<i>hasR</i>	121	CTGTCCGGTCTCTGAAGAGGGCACACCCGAATACCAGTACGCGCGCAACACGCTGATCGAG	180
<i>hasL</i>	121	CTGTCCGGTCTCTGAAGAGGGCACACCCGAATACCAGTACGCGCGCAACACGCTGATCGAG	180
<i>hasR</i>	181	ATGAACATGTCCCTGGTCCGCTTCGCGGCCGGCCGCTTCCGCAGCCGCGGACCGGAGGAG	240
<i>hasL</i>	181	ATGAACATGTCCCTGGTCCGCTTCGCGGCCGGCCGCTTCCGCAGCCGCGGACCGGAGGAG	240
<i>hasR</i>	241	ATGGAGGACATCGTCCAGGTCCGGCATGATCGGCCTGATCAAGGCCATCGACCGGTTTCGAG	300
<i>hasL</i>	241	ATGGAGGACATCGTCCAGGTCCGGCATGATCGGCCTGATCAAGGCCATCGACCGGTTTCGAG	300
<i>hasR</i>	301	CTCACCCGCGAGGTCGAGTTCACCTCCTTCGCCGTGCCCTACATCGTCCGGCAGATCAAG	360
<i>hasL</i>	301	CTCACCCGCGAGGTCGAGTTCACCTCCTTCGCCGTGCCCTACATCGTCCGGCAGATCAAG	360
<i>hasR</i>	361	CGGTTCTTCCGTGACACCTCCTGGGCCGTCCACGTGCCCGCGCGCTCCAGGAGGCCCGC	420
<i>hasL</i>	361	CGGTTCTTCCGTGACACCTCCTGGGCCGTCCACGTGCCCGCGCGCTCCAGGAGGCCCGC	420
<i>hasR</i>	421	GTCCAGCTGGCCCGGGCCACCGAGGAGCTGCGCAGCCGGCTGGGCGCAACCCACCACC	480
<i>hasL</i>	421	GTCCAGCTGGCCCGGGCCACCGAGGAGCTGCGCAGCCGGCTGGGCGCAACCCACCACC	480
<i>hasR</i>	481	AAGGAACTGTCCGAGCTGATGAGCCTGCCGGAGGACGAGGTCGTTCGAGGCCCGCTGGCG	540
<i>hasL</i>	481	AAGGAACTGTCCGAGCTGATGAGCCTGCCGGAGGACGAGGTCGTTCGAGGCCCGCTGGCG	540
<i>hasR</i>	541	GCCAACGGCTACAACCTCGCCTCCCTGGACGCGGCCATCAGCGGCAGCGAGGACGGCGAG	600
<i>hasL</i>	541	GCCAACGGCTACAACCTCGCCTCCCTGGACGCGGCCATCAGCGGCAGCGAGGACGGCGAG	600
<i>hasR</i>	601	TCCGCCCTGGCCGACTTCATCGGCGCCGACGACGCCACCTCGCCCTGGTGGACGACTTC	660
<i>hasL</i>	601	TCCGCCCTGGCCGACTTCATCGGCGCCGACGACGCCACCTCGCCCTGGTGGACGACTTC	660
<i>hasR</i>	661	CACGCGCTCGCCCCGATGATCGCCGAGCTCGACGAGCGCGACCGTTCAGATCATCCACTGG	720
<i>hasL</i>	661	CACGCGCTCGCCCCGATGATCGCCGAGCTCGACGAGCGCGACCGTTCAGATCATCCACTGG	720
<i>hasR</i>	721	CGGTTTCGTGGAGGAGCTCACCCAGAAGGACATCGGCGAGCGCTCGGCGTCTCCAGATG	780
<i>hasL</i>	721	CGGTTTCGTGGAGGAGCTCACCCAGAAGGACATCGGCGAGCGCTCGGCGTCTCCAGATG	780
<i>hasR</i>	781	CACGTCTCCCGGCTGATCTCGCGCCTGCTGGCCCGCTGCGCGAAGGCATGCTCACCGAC	840
<i>hasL</i>	781	CACGTCTCCCGGCTGATCTCGCGCCTGCTGGCCCGCTGCGCGAAGGCATGCTCACCGAC	840
<i>hasR</i>	841	CACTGA 846	
<i>hasL</i>	841	CTGA 846	

Figure R-15 : Alignement des séquences nucléotidiques des gènes *hasR* et *hasL*.  
 Les positions nucléotidiques divergentes sont sur fond vert lorsqu'elles sont spécifiques du gène *hasL*  
 et sur fond orange lorsqu'elles sont spécifiques du gène *hasR*.



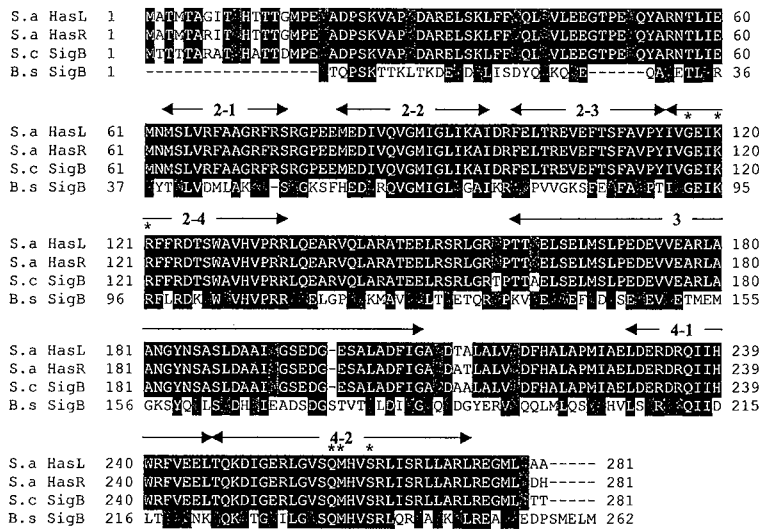


Figure R-17 : Alignement des séquences protéiques déduites des gènes *hasR* et *hasL* avec les séquences des facteurs  $\sigma^B$  de *S. coelicolor* et *B. subtilis*. Les différents domaines des facteurs sont indiqués. Les acides aminés, à l'intérieur des régions 2-4 et 4-2, impliqués dans la reconnaissance des boîtes -10 et -35 des promoteurs sont désignés par un astérisque.

	<i>S. a.</i> $\sigma^{HasR}$	<i>S. a.</i> $\sigma^{HasL}$
<i>S. a.</i> $\sigma^{HasL}$	97% (100% - 100%)	
<i>S. c.</i> $\sigma^B$ (SigJ)	93% (100% - 100%)	92% (100% - 100%)
<i>S. c.</i> SigL	86% (95.5% - 74%)	83% (95.5% - 74%)
<i>S. c.</i> CrtS	64% (100% - 62.5%)	63% (100% - 62.5%)
<i>S. c.</i> SigF	47% (73% - 62.5%)	46% (73% - 62.5%)
<i>B. s.</i> $\sigma^B$	34% (73% - 50%)	33% (73% - 50%)

Tableau R-6 : Homologues des facteurs sHas.

Pour chaque comparaison, le premier pourcentage représente la similarité entre les deux protéines. Les pourcentages d'identité entre les régions 2-4 et 4-2 respectivement sont donnés entre parenthèses.

La recherche de signaux de transcription et de traduction a été entreprise sur les séquences nucléotidiques.

Chez *Streptomyces*, la distance séparant le codon d'initiation de la traduction du site de fixation du ribosome (RBS pour Ribosome Binding Site) est généralement comprise entre 5 et 12 nucléotides (Strohl, 1992). Pour chacune des deux copies *hasR* et *hasL*, un RBS potentiel (GGAGG) a été identifié 10 pb en amont du codon ATG de début de traduction.

La recherche de séquences promotrices à partir de la seule connaissance de la localisation des signaux de traduction n'est possible que lorsqu'un consensus de promoteur a été défini.

Cependant, les connaissances sur les signaux de transcription chez les *Streptomyces* sont extrêmement parcellaires. Elles dérivent de l'identification de motifs nucléotidiques hautement représentés en amont des sites d'initiation de la transcription (+1) pour 150 gènes issus de différentes espèces de *Streptomyces* (Bourn and Babb, 1995). Des motifs consensus ont été identifiés en position -35 et -10, organisation typique des promoteurs dépendants de  $\sigma 70$  d'*E. coli*, dans 20% des gènes seulement. Ces motifs ont été recherchés en amont des gènes *has* et plus particulièrement en amont des sites d'initiation de transcription identifiés (voir paragraphe B-III – 2.1.2).

Chez *E. coli*, un terminateur de transcription  $\rho$ -indépendant peut être prédit par l'identification d'une structure tige-boucle possédant une région riche en bases G+C au pied de la tige. Chez les *Streptomyces*, compte tenu de la richesse en bases G+C de l'ADN génomique, la présence de structures tiges-boucles est fréquente. Ainsi, la composition en bases de la tige ne peut être indicative du caractère fonctionnel de cette structure comme site d'arrêt de la transcription. La présence de structures secondaires en aval du codon d'arrêt de la traduction des gènes *has* a été recherchée à l'aide du logiciel DNA mfold.

La présence d'une structure secondaire correspondant à un terminateur de transcription potentiel a ainsi été détectée 29 pb en aval du codon d'arrêt de traduction de *hasR*. Cette structure correspond à un palindrome imparfait de 29 pb (tige de 14 pb interrompue par 4 mésappariements, boucle de 1 nucléotide). De façon similaire, la recherche de structure secondaire en aval de *hasL* a révélé la présence d'une tige-boucle (tige de 33 pb avec 8 mésappariements, boucle de 4 nucléotides). Cette structure est localisée 220 pb en aval du codon d'arrêt de traduction et pourrait servir de terminateur de transcription. Une structure de même type, mais de moindre stabilité, a également été détectée : palindrome imparfait de 30 pb (tige de 12 pb interrompue par 3 mésappariements, boucle de 6 nucléotides) localisé à 30 pb en aval du codon d'arrêt de la traduction de *hasL*.

### **B-I.1.2 – L'ENVIRONNEMENT GENETIQUE DES GENES HAS**

Afin de déterminer l'environnement génétique des gènes *has*, le séquençage des régions amont et aval de chacun des deux gènes a été entrepris.

Le séquençage des régions amont des gènes *hasL* et *hasR* a été réalisé sur l'ADN des inserts des bactériophages M13 recombinants pl329 et pl338 respectivement et la région aval du gène *hasR* sur l'ADN de l'insert du bactériophage M13 recombinant pl73. L'ADN du cosmide recombinant 14C4 a quant à lui servi de matrice au séquençage de la région aval du gène *hasL*.

Ainsi, un total de 4429 pb pour *hasR* et 3659 pb pour *hasL* a été séquencé.

L'analyse de la séquence du locus *hasR* par le logiciel Frameplot 3.0beta a permis de révéler la présence 5 ORF appelées *orfA*, *arsR*, *prsR*, *hasR* et *orfB* (Figure R-18).

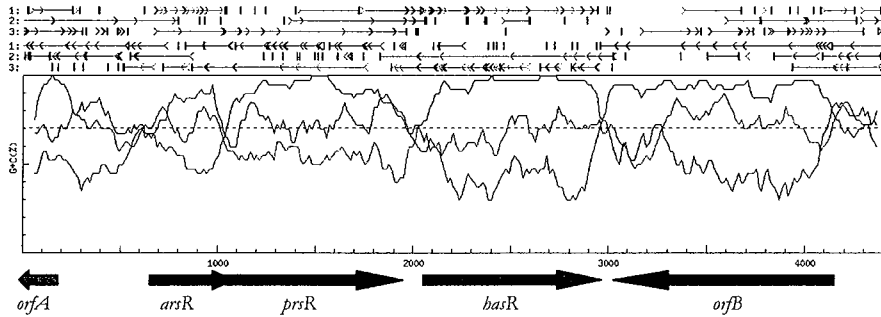


Figure R-18 : Analyse des séquences du locus *hasR*.

A. Résultat de l'analyse fourni par le logiciel Frameplot 3.0beta.

Sur la partie supérieure sont présentés les signaux de traduction. Un trait vertical correspond à un codon stop (UAA, UAG, UGA). Une flèche indique un codon start (ATG, GTG). Un trait relie un codon start et un codon stop lorsque l'ORF est d'une longueur minimale de 20 acides aminés. 1, 2, 3 et -1, -2, -3 correspondent aux six phases de lecture de la séquence nucléotidique.

La partie inférieure schématise le pourcentage en bases G+C aux trois positions des codons. Le trait pointillé matérialise le pourcentage moyen en bases G+C de la séquence.

B. Représentation schématique de l'organisation du locus *hasR*. Le code couleur utilisé est identique à celui de la partie A.

La protéine codée par le gène *prsR* présente 89% de similarité avec l'anti- $\sigma$  putatif du facteur  $\sigma^B$  chez *S. coelicolor*. La recherche de domaines protéiques a révélé la présence d'un domaine RsbW (COG2172), caractéristique des anti-facteurs  $\sigma$ , ainsi que celle d'un domaine HATPase-c (pfam02518), suggérant l'activité kinase putative de cet anti-facteur  $\sigma$  (Figure R-19).

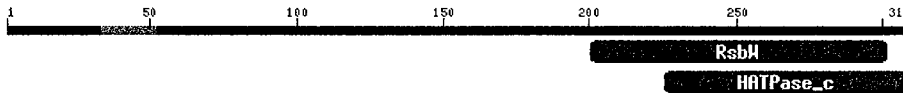


Figure R-19 : Localisation des domaines RsbW (COG2172) et HATPase-c (pfam02518) sur la séquence de l'anti-facteur  $\sigma$  PrsR.

Les positions des acides aminés sont indiquées en bleu. Les différents domaines protéiques sont symbolisés par les rectangles rouges.

Une analyse similaire sur le gène *arsR* a révélé que la protéine putative codée par ce gène présentait 91% de similarité avec l'anti-anti-facteur  $\sigma^B$  de *S. coelicolor*.

De plus, deux domaines caractéristiques des anti-anti-facteurs  $\sigma$  bactériens ont pu être identifiés sur la séquence de cette protéine putative : le domaine SpoIIAA (COG1366) et le domaine STAS (pfam01740) (Figure R-20).



Figure R-20 : Localisation des domaines SpoIIAA (COG1366) et STAS (pfam01740) sur la séquence de l'anti-anti-facteur  $\sigma$  ArsR.

Le symbolisme est identique à celui de la figure R-19.

Il est intéressant de noter que le codon d'initiation de la traduction du gène *prsR* est chevauchant avec le codon d'arrêt de la traduction du gène *arsR* (séquence **ATGA**) suggérant que ces deux gènes pourraient être cotranscrits. En accord avec ces données, les analyses transcriptionnelles des gènes *has* ont en effet révélé que le gène *hasR* faisait partie d'une structure opéronique englobant les gènes *arsR/prsR/hasR* (voir paragraphe B-III.2.1).

Les deux autres ORF, appelées *orfA* et *orfB*, coderaient respectivement une protéine hypothétique (*orfA* partiellement séquencée, 85% d'identité et 86% de similarité avec les 104 premiers acides aminés de la protéine hypothétique conservée codée par l'ORF SCF55-21c de *S. coelicolor*) et une protéine sécrétée (70% d'identité et 77% de similarité avec la protéine putative sécrétée codée par l'ORF SCF55-25c de *S. coelicolor*).

La comparaison de l'ensemble de ces séquences avec les banques de données du génome de *S. coelicolor* a permis de révéler une situation originale : en effet l'ensemble de ces gènes ainsi que leur organisation sont apparus conservés chez *S. coelicolor* autour du gène *sigB* (Cho *et al.*, 2001), homologue des gènes *has*, mais leur localisation chromosomique, sur le bras gauche du chromosome de *S. coelicolor*, semble révéler un grand remaniement entre les deux espèces *S. coelicolor* et *S. ambofaciens*, puisque le locus *hasR* est quant à lui localisé sur le bras chromosomique droit.

Afin de préciser l'ampleur du réarrangement ainsi mis en évidence des expériences d'hybridation de cosmides recombinants correspondant aux régions flanquant les gènes *has* sur des profils de PFGE de l'ADN total de *S. coelicolor* et *S. ambofaciens* ont été réalisées. Ainsi, les trois cosmides recombinants 8C4, 25E1 et 6E4 recouvrant environ 160 kb autour du locus *hasR* (Figure R-21) ont été utilisés comme sondes. Ces trois cosmides, cartographiés dans le fragment *AseI*-E localisé sur le bras chromosomique droit de *S. ambofaciens*, révèlent le fragment *AseI*-F sur le profil d'ADN de *S. coelicolor* correspondant à l'extrémité gauche du chromosome (résultats non montrés).

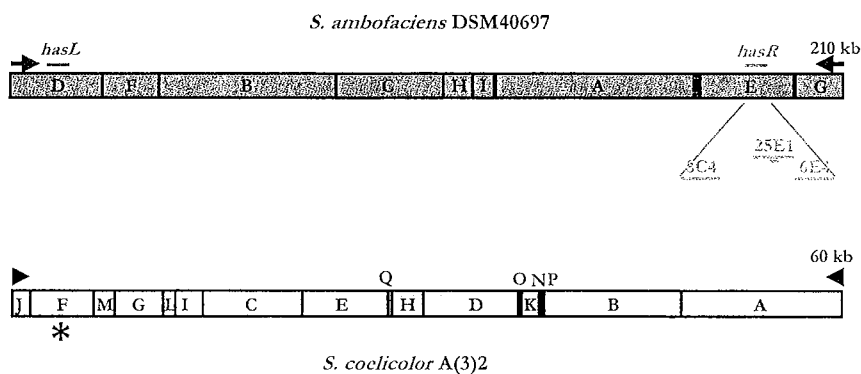


Figure R-21 : Localisation des cosmides recombinants 8C4, 25E1 et 6E4 dans le fragment *AseI* E du chromosome de *S. ambofaciens* et localisation du fragment *AseI* F, révélé par ces cosmides sur le chromosome de *S. coelicolor*.



Ainsi, alors que les cartes génétiques des chromosomes de *S. ambofaciens* et *S. coelicolor* sont collinéaires pour les gènes au centre du chromosome, le cas du locus *hasR* confirme la grande variabilité des régions terminales (Figure R-22), à l'image de la localisation différente des gènes *spa2* et *spaR*, codant des régulateurs transcriptionnels d'expression en phase tardive mais également de la taille différente des répétitions terminales inversées (60 kb chez *S. coelicolor* contre 210 kb chez *S. ambofaciens*)

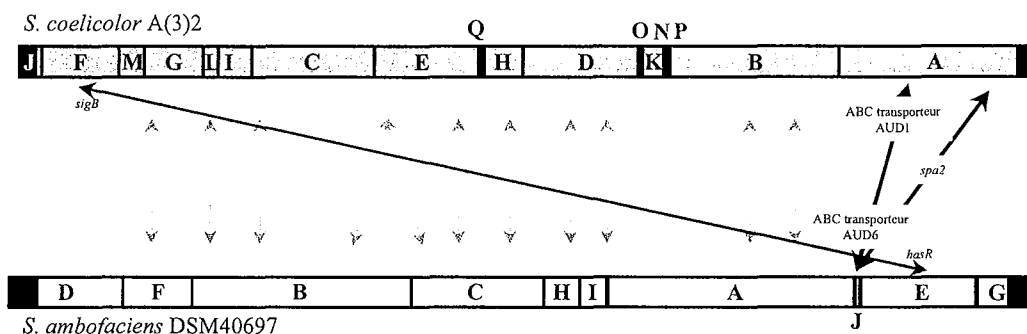


Figure R-22 : Cartographie comparée de *S. ambofaciens* DSM40697 et *S. coelicolor* A3(2). Selon Leblond *et al.* 1996. Cette comparaison révèle une bonne conservation de l'ordre des gènes entre les deux espèces au centre du chromosome et une faible conservation dans les régions terminales.

En ce qui concerne le locus *hasL*, deux ORF ont pu être identifiées suite au séquençage (Figure R-23).

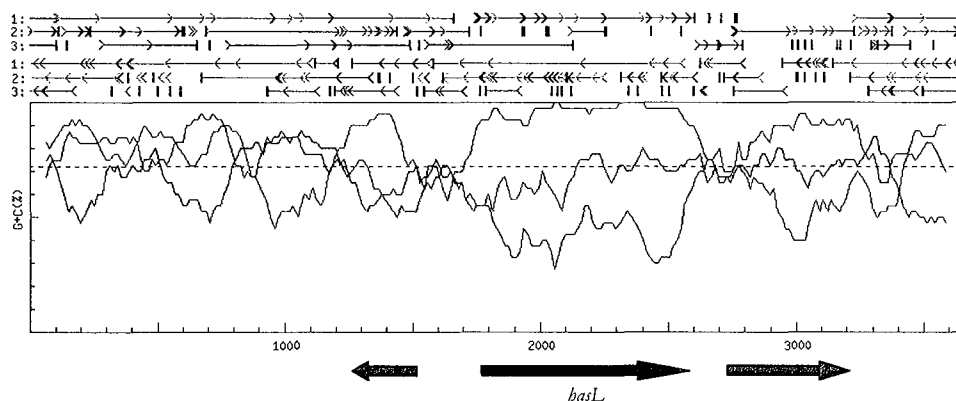


Figure R-23 : Analyse des séquences du locus *hasL*. Le symbolisme est identique à celui utilisé pour la Figure R-18.

La première ORF, nommée *orfC*, présente 78% d'identité et 90% de similarité avec la protéine hypothétique conservée SC5E9-03 de *S. coelicolor*. La seconde ORF a été appelée *prsL* car coderait une protéine régulatrice de type anti-facteur  $\sigma$  avec les domaines protéiques HATase-c et RsbW

(Figure R-24). Il est cependant à noter que ces domaines sont bien moins conservés dans le cas de *prsL*, comparativement à *prsR* (selon les valeurs « e values » données par le logiciel : <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

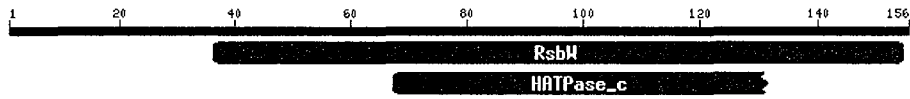


Figure R-24 : Localisation des domaines RsbW (COG2172) et HATPase-c (pfam02518) sur la séquence de l'anti-facteur  $\sigma$  PrsL.

Le symbolisme est identique à celui de la figure R-19.

L'homologue de *prsL* chez *S. coelicolor* (43% d'identité protéique), nommé SC5H1-15c est retrouvé en aval du gène *sigL*, codant un facteur  $\sigma$  présentant 71% d'identité avec  $\sigma^{\text{HasR}}$ .

## B-I.2 Les gènes chimériques *HAS135* et *HAS855* chez les mutants *Streptomyces ambofaciens* NSAH135 et NSA855

Le réarrangement chromosomique présent chez les mutants *Streptomyces ambofaciens* NSAH135 et NSA855 (Fischer *et al.*, 1998b) consiste en un événement de recombinaison homologue ectopique entre les séquences *hasR* et *hasL*. Outre la délétion et la duplication de nombreux gènes, la conséquence de ces échanges d'extrémités chromosomiques est la formation d'ORF chimériques au point de recombinaison.

La souche mutante NSAH135 renferme une ORF chimérique *has135* composée de la région 5' de *hasL* et de la région 3' de *hasR* et réciproquement pour la souche mutante NSA855 (Figure R-25). Les translocations chromosomiques entraînent donc l'apparition d'un cadre de régulation chimérique pour ces nouveaux gènes *has* associant des signaux 5' spécifiques d'une copie *has* aux signaux de régulation en 3' de l'autre copie *has*, comme par exemple le signal hypothétique de dégradation par les protéases spécifique de *hasL*.

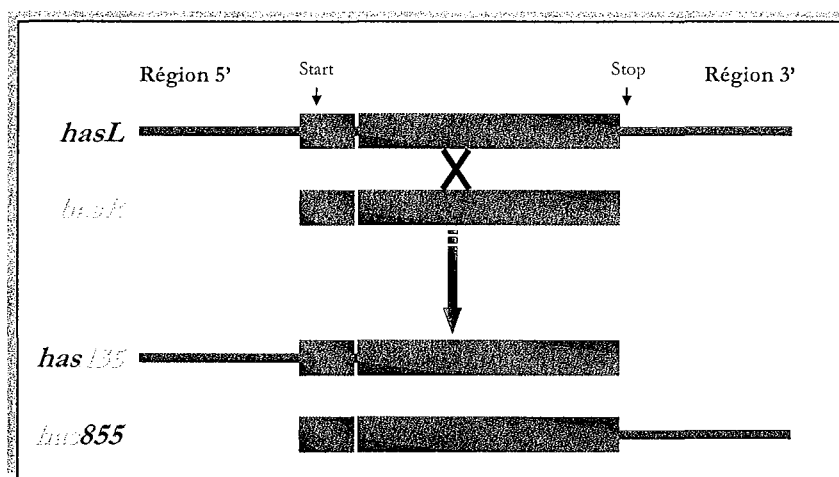
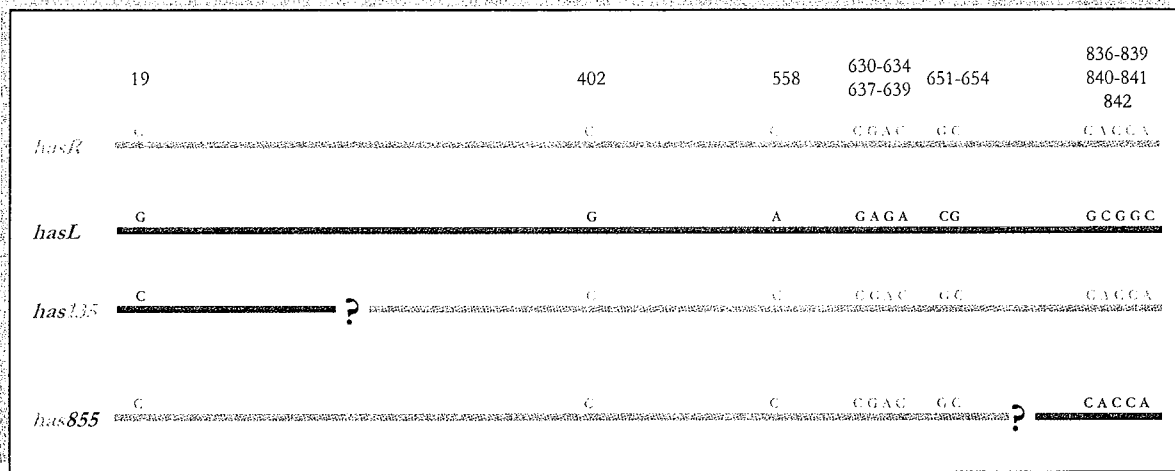


Figure R-25 : Représentation schématique de la formation de chimères des ORF *has* chez les souches mutantes *S. ambofaciens* NSAH135 et NSA855.

La croix représente l'événement de recombinaison homologue. La zone d'identité nucléotidique à 100% entre les loci *has* est représentée par un rectangle bleu. Start et Stop correspondent respectivement aux codons d'initiation et d'arrêt de la traduction des ORF. Les régions 5' et 3' correspondent aux régions situées en amont et en aval des ORF.

Il est important de noter que les souches NSAH135 et NSA855 ont été isolées de manière indépendante, et que d'autre part, l'analyse des séquences des gènes chimériques *has135* et *has855* permet de conclure quant à une localisation distincte du point de recombinaison dans les deux cas (pour NSA855 entre les positions 654 et 836, chez NSAH135PP entre les positions 19 et 402, Figure R-26).



**Figure R-26 : Mise en évidence d'ORF chimériques issues d'événements de recombinaison homologue distincts entre les gènes *hasR* et *hasL*.**

En haut, positions nucléotidiques. En dessous, schématisation de l'alignement des ORF sauvages et mutantes où seules les positions divergentes sont indiquées. Le point d'interrogation présent dans les ORF chimériques *has135* et *has855* indique que la position exacte de l'événement de recombinaison homologue n'est pas connue.

- CHAPITRE B-II -

LA DUPLICATION DES GENES *HAS*  
CHEZ *STREPTOMYCES AMBOFACIENS*  
EST RECENTE

Afin de déterminer si la duplication des gènes *has* chez *S. ambofaciens* est un évènement récent ou non, le statut (dupliqué ou non) du (ou des) homologue(s) des gènes *has* a été recherché chez d'autres espèces de *Streptomyces*. Préalablement à cette analyse, la conservation relative de l'architecture chromosomique a été démontrée au niveau intraspécifique chez plusieurs souches de *S. ambofaciens* grâce au séquençage des séquences ITS. Cette analyse a également permis de confirmer les liens phylogénétiques entre les espèces *S. ambofaciens* et *S. coelicolor*.

## B-II.1 $\sphericalangle$ CONSERVATION DE L'ARCHITECTURE DE LA REGION « CORE » DU CHROMOSOME DE *S. AMBOFACIENS*

*Les résultats résumés dans ce chapitre sont exposés dans la publication (Wenner et al., 2002) (page 329).*

Afin d'estimer la part des réarrangements chromosomiques dans l'évolution de la région « core » du chromosome des *Streptomyces* et de confirmer les liens phylogénétiques des différents *Streptomyces*, les séquences des ITS de deux souches de *S. ambofaciens* et de deux autres espèces, *S. coelicolor* M145 et *S. griseus* 2247, ont été comparées.

### B-II.1.1 – DETERMINATION DES SEQUENCES DES ITS DE *S. AMBOFACIENS*

Chez la plupart des *Streptomyces*, 6 loci *rrn* (codant les ARN ribosomiques) ont été mis en évidence (Cole and Saint-Girons, 1999). Ils sont organisés en opéron, 16S-23S-5S (Figure R-27). Chez de nombreuses bactéries, ces loci *rrn* sont le substrat de fréquents réarrangements chromosomiques, allant du remplacement de petits blocs de nucléotides jusqu'à de grands réarrangements, pouvant même faire varier le nombre de chromosomes au sein d'une même espèce (Jumas-Bilak *et al.*, 1998). Ces loci *rrn* peuvent aussi constituer le substrat de nombreuses études phylogénétiques. Une séquence située entre les gènes 16S et 23S sert notamment à établir des relations phylogénétiques entre espèces proches. Cette séquence est appelée séquence interne transcrite ou ITS (Internally Transcribed Sequence, Figure R-27).



Figure R-27 : Organisation classique d'un locus codant pour l'ARN ribosomique chez les bactéries.

Les oligonucléotides universels R et L servent à amplifier l'ITS 16S-23S.

Afin de déterminer les séquences des ITS de *S. ambofaciens* DSM40697 et ATCC23877, les 6 loci *rrn* A à F de chaque souche ont été isolés dans des cosmides recombinants dont l'ADN a ensuite servi de matrice pour l'amplification des ITS lors d'une PCR avec les oligonucléotides universels L et R (Figure R-27). Un produit de PCR d'environ 300 pb a été obtenu pour les 12 loci et chaque produit de PCR a été séquencé.

### B-II.1.2 - MISE EN EVIDENCE DE BLOCS CONSTANTS ET VARIABLES DANS LES ITS DE STREPTOMYCES

L'analyse des séquences des ITS présentes dans chaque souche montre qu'elles se composent de 5 blocs de nucléotides (c1 à c5) identiques entre les six ITS de chaque souche (Figure R-28). Ces blocs constants sont séparés par 4 blocs de séquences variables (v1 à v4, Figure R-28). Ces blocs de séquences variables sont composés de 3 séquences différentes pour les blocs variables v1-v3 et de deux pour le bloc variable v4. Les types de blocs partagent entre eux moins de 60% d'identité nucléotidique. En tout, 5 combinaisons de blocs variables peuvent être distinguées dans chaque souche (Tableau R-7).

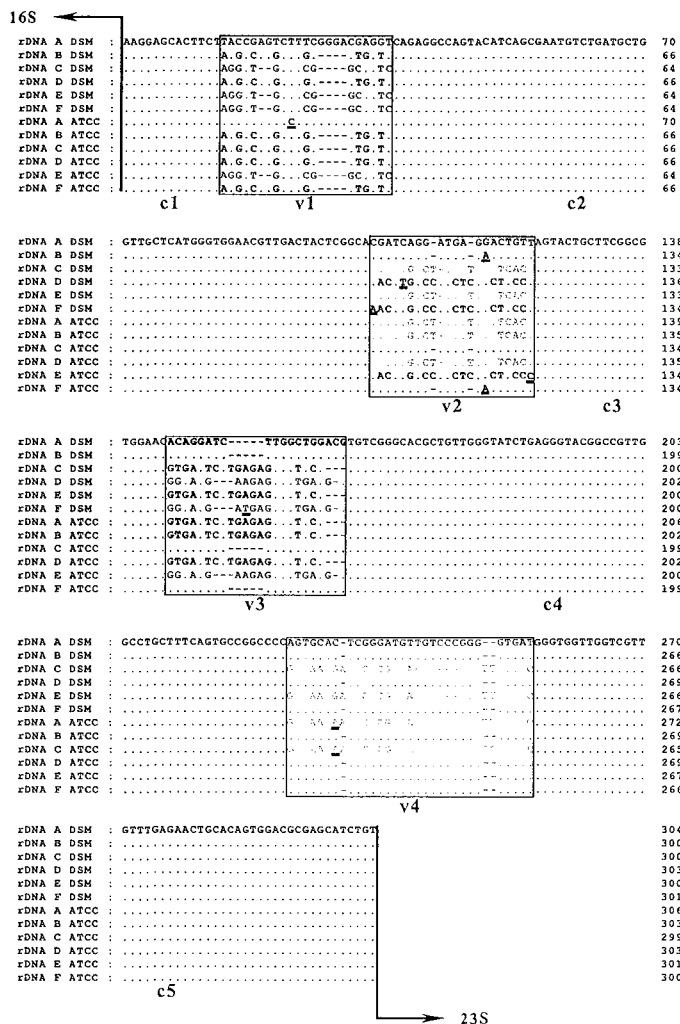


Figure R-28 : Comparaison des 6 séquences des ITS de *S. ambofaciens* DSM40697 (rDNA A-F DSM) avec les 6 séquences des ITS de *S. ambofaciens* ATCC 23877 (rDNA A-F ATCC). Alignement des douze séquences montrant les 5 régions constantes (c1 à c5) et les 4 régions variables (v1 à v4, encadrées). Les différences entre régions variables du même type sont soulignées.

**B-II.1.3 – COMPARAISON DES SEQUENCES ITS**

Au niveau intraspécifique, la comparaison des séquences ITS entre les souches DSM40697 et ATCC23877 montre que ces souches partagent les mêmes types de séquences variables, aucune séquence variable n'étant spécifique d'une souche. De plus, les paires de loci homologues ne présentent jamais la même combinaison de blocs variables et des loci non homologues peuvent partager la même organisation. (Tableau R-7). L'ensemble de ces résultats suggère donc fortement que les loci *rrn* de *Streptomyces ambofaciens* sont le siège d'événements de recombinaison homologue conduisant à des échanges de blocs de nucléotides.

v1	v2	v3	v4	
				iDNA A DSM
v1.1				iDNA A ATCC
v1.2				iDNA B DSM iDNA F ATCC
				iDNA B ATCC iDNA D ATCC
				iDNA D DSM
				iDNA C DSM iDNA E DSM
v1.3				iDNA F DSM iDNA E ATCC

Tableau R-7 : Association des blocs variables dans les loci *rrn*.

Afin de réaliser une comparaison interspécifique de ces régions, les 6 séquences ITS des souches *S. coelicolor* M145 et sur *S. griseus* 2247, disponibles dans les banques de données, ont été analysées (Figure R-29). Cette comparaison a été effectuée sur 5 loci sur 6 chez *S. griseus*, la séquence du dernier locus divergeant fortement suggérant que cet ITS, voire l'ensemble du locus *rrn* ait été acquis par transfert horizontal.

Le même type de région modulaire a ainsi été caractérisé chez ces deux espèces, bien que le nombre de régions variables soit plus limité.

La comparaison des régions constantes montre l'existence d'une région variable entre espèces dans la région constante c4. Outre cette région, l'analyse des régions constantes montre l'existence de 9 positions nucléotidiques divergentes entre *S. griseus* et l'ensemble *S. ambofaciens*/*S. coelicolor* (Figure R-29).

L'analyse des séquences montre l'existence de 4 blocs de séquences variables spécifiques de *S. griseus* et révèle également que deux types de séquences variables supplémentaires sont présentes chez *S. coelicolor* par rapport à *S. ambofaciens*. Ainsi, la structure modulaire de *S. ambofaciens* est retrouvée chez *S. coelicolor* mais pas chez *S. griseus*.



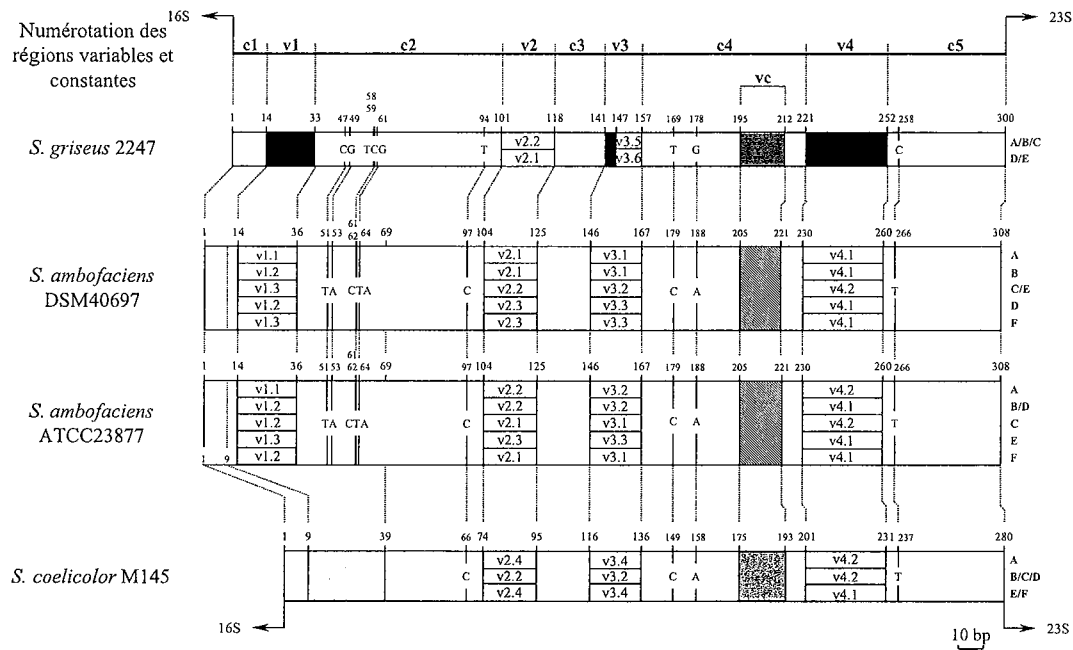


Figure R-29 : Comparaison des ITS de *S. ambofaciens* (DSM40697 et ATCC23877), *S. coelicolor* M145 et *S. griseus* 2247.

Les régions spécifiques de *S. griseus* sont en noires et celles spécifiques de *S. coelicolor* sont en gris. Les régions variables entre espèces mais constantes entre souches de la même espèce sont désignées par vc et sont indiquées par des lignes hachurées horizontales ou obliques respectivement chez *S. griseus* ou *S. ambofaciens* et par un rectangle gris hachuré chez *S. coelicolor*. La numérotation dérive de l'alignement Figure 2R. Les différences entre les régions constantes sont indiquées alors que celles entre régions variables ne le sont pas.

#### B-II.1.4 - PREDICTION DES STRUCTURES SECONDAIRES DU REPLIEMENT DES SEQUENCES ITS

Un modèle du repliement entre les ITS 16S-23S et 23S-5S peut enfin être proposé (Figure R-30). Il est intéressant de noter que les différences nucléotidiques observées entre les régions constantes de *S. griseus* et le groupe *S. ambofaciens/S. coelicolor* (Figure R-29) correspondent à des mutations compensatoires conduisant au maintien d'une structure secondaire analogue.



## **B-II.2 $\omega$ CONSERVATION DE LA DUPLICATION DES GENES *HAS* CHEZ LES TROIS SOUCHES DE *STREPTOMYCES AMBOFACIENS***

Le séquençage des gènes *has* a également été entrepris chez les trois souches de *S. ambofaciens* DSM40697, ATCC23877 et ETH12567.

Ainsi, les gènes *hasR* et *hasL* ont été spécifiquement amplifiés par PCR à l'aide d'amorces localisées dans les régions divergentes entre les deux gènes. Puis chaque produit de PCR a été séquencé à l'aide des amorces ayant servi aux réactions de PCR.

Il est ainsi apparu la duplication des gènes *has* était présente chez les trois souches de *S. ambofaciens*, puisque les deux copies *hasR* et *hasL* présentent 97% d'identité nucléotidique chez les souches ATCC23877 comme ETH12567, comparativement aux 98% observés chez la souche DSM40697 (Figure R-31).



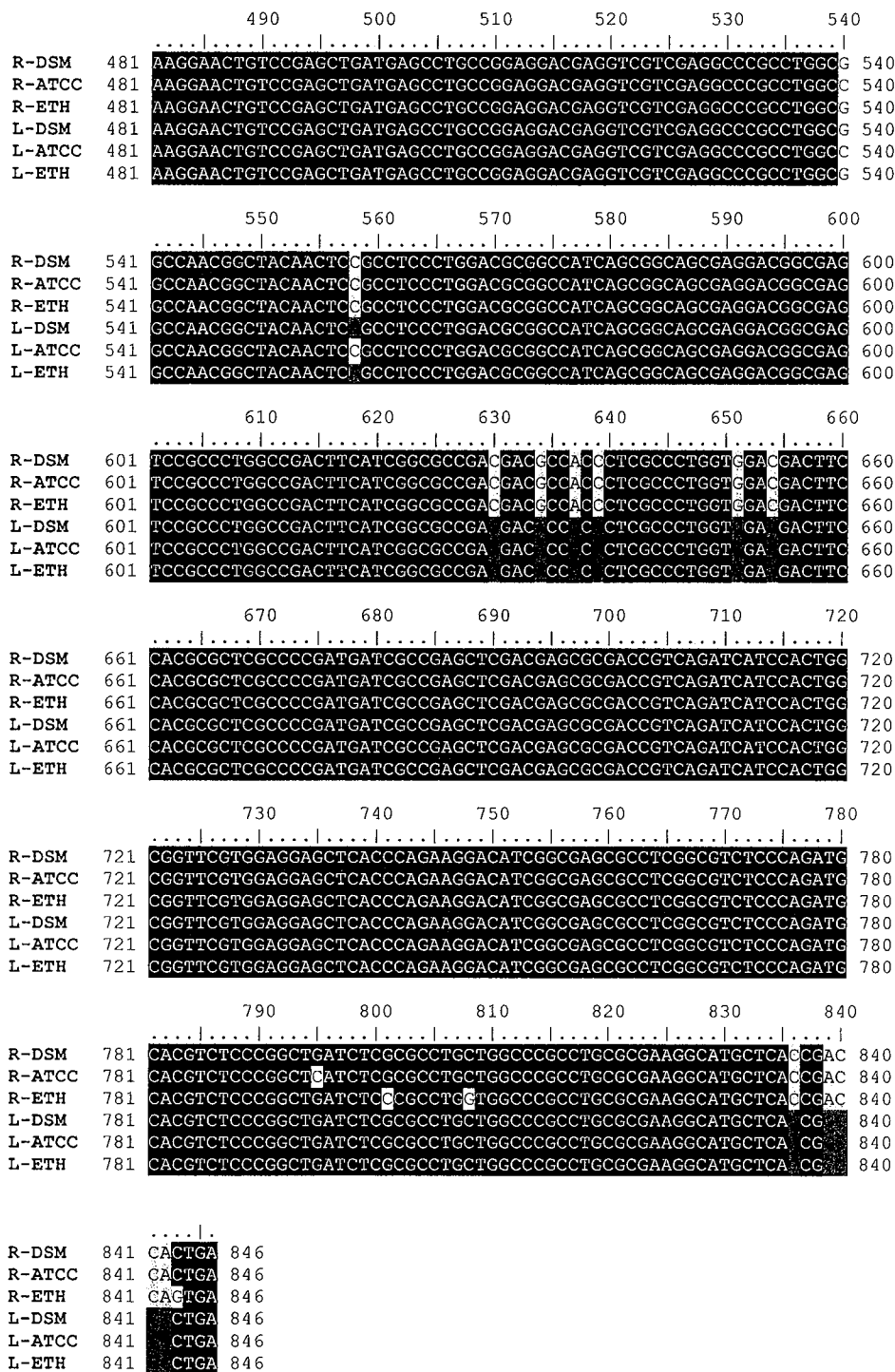


Figure R-31 : Alignement des séquences nucléotidiques des gènes *hasR* (R) et *hasL* (L) chez les souches *S. ambofaciens* DSM40697 (DSM), ATCC23877 (ATCC) et ETH12567 (ETH). Les positions nucléotidiques divergentes sont sur fond vert lorsqu'elles sont spécifiques du gène *hasL* et sur fond orange lorsqu'elles sont spécifiques du gène *hasR*.

De plus, la majorité des mutations nucléotidiques observées sont de type conservatives et ne changent pas la séquence en acides aminés ou la changent en acides aminés similaires (Figure R-32). Enfin, il est à noter que les différences en acides aminés entre les deux facteurs  $\sigma^{\text{Has}}$  observées chez la souche DSM40697 sont également retrouvées chez les deux autres souches de *S. ambofaciens*, à l'exception de la position 281 de la protéine HasR chez la souche ETH12567, suggérant l'existence d'une pression de sélection pour la conservation de ces acides aminés spécifiques et donc pour la conservation de la duplication (voir Discussion).

		10	20	30	40	50	60	
R-DSM	1	MATMTAR	I	TEHTTTGMPEVADPSKVAPKDARELSK	LF	FQQLSVLEEGTPEYQYARNTL	IE	60
R-ATCC	1	MATMTAR	I	TEHTTTGMPEVADPSKVAPKDARELSK	LF	FQQLSVLEEGTPEYQYARNTL	IE	60
R-ETH	1	MATMTAR	I	TEHTTTGMPEVADPSKVAPKDARELSK	LF	FQQLSVLEEGTPEYQYARNTL	IE	60
L-DSM	1	MATMTA	I	TEHTTTGMPEVADPSKVAPKDARELSK	LF	FQQLSVLEEGTPEYQYARNTL	IE	60
L-ATCC	1	MATMTA	I	TEHTTTGMPEVADPSKVAPKDARELSK	LF	FQQLSVLEEGTPEYQYARNTL	IE	60
L-ETH	1	MATMTA	I	TEHTTTGMPEVADPSKVAPKDARELSK	LF	FQQLSVLEEGTPEYQYARNTL	IE	60
		70	80	90	100	110	120	
R-DSM	61	MNMSLVRFAAGRFRSRGPEEMEDI	VQVGMIGLIKAI	DRFEL	TREVEFTS	FAVPY	IVGEIK	120
R-ATCC	61	MNMSLVRFAAGRFRSRGPEEMEDI	VQVGMIGLIKAI	DRFEL	TREVEFTS	FAVPY	IVGEIK	120
R-ETH	61	MNMSLVRFAAGRFRSRGPEEMEDI	VQVGMIGLIKAI	DRFEL	TREVEFTS	FAVPY	IVGEIK	120
L-DSM	61	MNMSLVRFAAGRFRSRGPEEMEDI	VQVGMIGLIKAI	DRFEL	TREVEFTS	FAVPY	IVGEIK	120
L-ATCC	61	MNMSLVRFAAGRFRSRGPEEMEDI	VQVGMIGLIKAI	DRFEL	TREVEFTS	FAVPY	IVGEIK	120
L-ETH	61	MNMSLVRFAAGRFRSRGPEEMEDI	VQVGMIGLIKAI	DRFEL	TREVEFTS	FAVPY	IVGEIK	120
		130	140	150	160	170	180	
R-DSM	121	RFFRDTSWAVHVP	RRRLQEARVQLARATEELRSRLGRNPTTKELSELMSLPEDEV	VEARLA				180
R-ATCC	121	RFFRDTSWAVHVP	RRRLQEARVQLARATEELRSRLGRNPTTKELSELMSLPEDEV	VEARLA				180
R-ETH	121	RFFRDTSWAVHVP	RRRLQEARVQLARATEELRSRLGRNPTTKELSELMSLPEDEV	VEARLA				180
L-DSM	121	RFFRDTSWAVHVP	RRRLQEARVQLARATEELRSRLGRNPTTKELSELMSLPEDEV	VEARLA				180
L-ATCC	121	RFFRDTSWAVHVP	RRRLQEARVQLARATEELRSRLGRNPTTKELSELMSLPEDEV	VEARLA				180
L-ETH	121	RFFRDTSWAVHVP	RRRLQEARVQLARATEELRSRLGRNPTTKELSELMSLPEDEV	VEARLA				180
		190	200	210	220	230	240	
R-DSM	181	ANGYNSASLDAAISGSE	DGESALADFIGA	D	ATLALV	DFHALAPMIAELDERDROI	IHW	240
R-ATCC	181	ANGYNSASLDAAISGSE	DGESALADFIGA	D	ATLALV	DFHALAPMIAELDERDROI	IHW	240
R-ETH	181	ANGYNSASLDAAISGSE	DGESALADFIGA	D	ATLALV	DFHALAPMIAELDERDROI	IHW	240
L-DSM	181	ANGYNSASLDAAISGSE	DGESALADFIGA	D	LALV	DFHALAPMIAELDERDROI	IHW	240
L-ATCC	181	ANGYNSASLDAAISGSE	DGESALADFIGA	D	LALV	DFHALAPMIAELDERDROI	IHW	240
L-ETH	181	ANGYNSASLDAAISGSE	DGESALADFIGA	D	LALV	DFHALAPMIAELDERDROI	IHW	240
		250	260	270	280			
R-DSM	241	RFVEELTQKDIGERL	GVSQMHVSRLISRL	LAR	REGML	DH	281	
R-ATCC	241	RFVEELTQKDIGERL	GVSQMHVSRLISRL	LAR	REGML	DH	281	
R-ETH	241	RFVEELTQKDIGERL	GVSQMHVSRLISRL	AR	REGML	DQ	281	
L-DSM	241	RFVEELTQKDIGERL	GVSQMHVSRLISRL	LAR	REGML		281	
L-ATCC	241	RFVEELTQKDIGERL	GVSQMHVSRLISRL	LAR	REGML		281	
L-ETH	241	RFVEELTQKDIGERL	GVSQMHVSRLISRL	LAR	REGML		281	

Figure R-32 : Alignement des séquences protéiques déduites des gènes *hasR* (R) et *hasL* (L) chez les souches *S. ambofaciens* DSM40697 (DSM), ATCC23877 (ATCC) et ETH12567 (ETH).

Le symbolisme est identique à celui de la Figure R-31.

### B-II.3 LES GENES *HAS* AU SEIN DU GENRE *STREPTOMYCES*

La récente disponibilité de la séquence du génome de *S. avermitilis* (Ikeda *et al.*, 2003) avec l'analyse concertée de la séquence du génome de *S. coelicolor* (Bentley *et al.*, 2002) nous a permis d'inspecter l'état, dupliqué ou non, des orthologues des gènes *has* chez ces deux espèces.

Chez *S. coelicolor* comme chez *S. avermitilis*, une famille de 9 gènes, définis comme codant des homologues du facteur  $\sigma^B$ , doit être considérée.

Or, il apparaît que l'orthologue direct des gènes *has* appartient à cette famille. Il s'agit du gène :

- *sigJ* chez *S. coelicolor*, encore appelé *sigB* selon Cho *et al.* 2001), dont le produit présente 93% de similarité avec  $\sigma^{HasR}$ ;
- SAV741 chez *S. avermitilis*, codant un facteur  $\sigma$  présentant 80% de similarité avec  $\sigma^{HasR}$ .

Les gènes présentant la seconde plus forte identité avec les gènes *has* dans ces familles sont les gènes *sigL* chez *S. coelicolor* et SAV1151 chez *S. avermitilis* dont les produits présentent tous deux 71% de similarité avec  $\sigma^{HasR}$ .

L'alignement des séquences protéiques révèle que les couples SigJ/SigL pour *S. coelicolor* et SAV741/SAV1151 pour *S. avermitilis* présentent une identité de 66%. Comparativement à la l'homologie de 98% retrouvée au niveau nucléotidique entre les gènes *hasR* et *hasL* chez *S. ambofaciens*, il apparaît donc que la duplication des gènes *has* n'est pas retrouvée chez ces deux espèces.

Compte-tenu des relations phylogénétiques liant ces trois espèces (Figure R-33), la duplication des gènes *has* chez *S. ambofaciens* serait un événement récent au niveau de l'évolution.

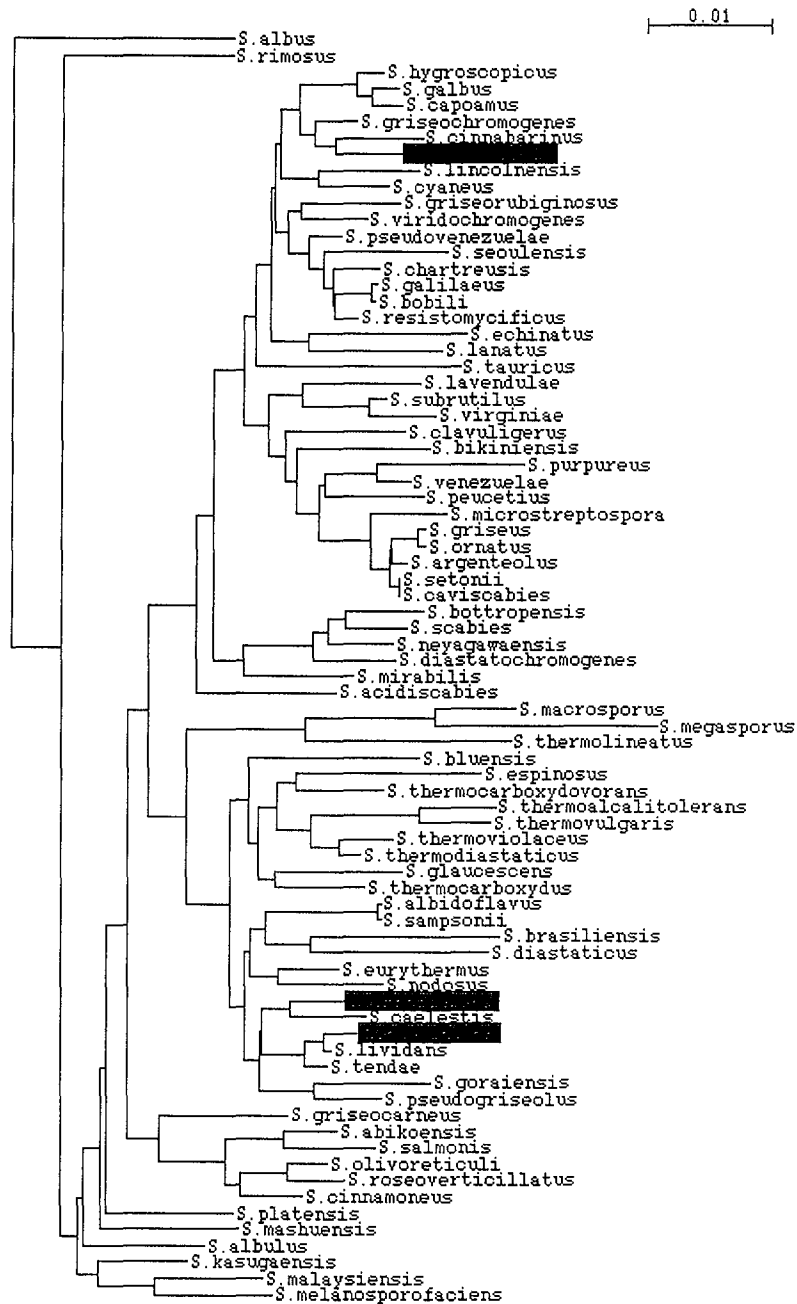


Figure R-33: Arbre phylogénétique des espèces de *Streptomyces* construit à partir des séquences des ARNr 16S.

Selon Ikeda *et al.* (2003) (<http://avermitilis.ls.kitasato-u.ac.jp/>).

De plus, lorsque un arbre phylogénétique est dessiné avec les séquences des gènes *hasR* et *hasL* ainsi qu'avec celles des 9 gènes de *S. coelicolor* et de *S. avermitilis*, deux sous familles peuvent notamment être distinguées (Figure R-34) :

□ un premier groupe renfermant le gène *sigJ* de *S. coelicolor*, le gène SAV741 de *S. avermitilis* ainsi que les deux copies *hasR* et *hasL* de *S. ambofaciens* ;



□ et un second groupe constitué des gènes *sigL* de *S. coelicolor* et SAV1151 de *S. avermitilis*.

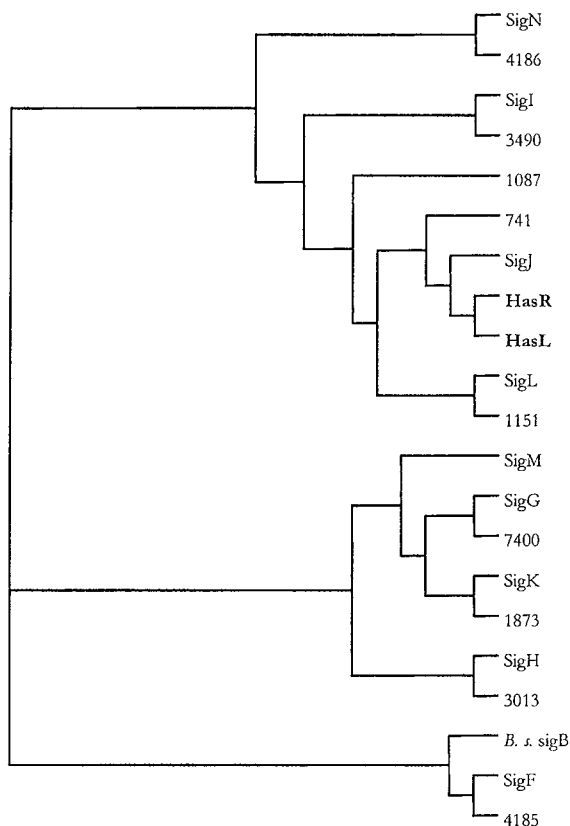


Figure R-34 : Arbre phylogénétique des gènes *hasR* et *hasL* ainsi que de leurs homologues chez *S. coelicolor* et *S. avermitilis*.

Cet arbre a été construit par la méthode des distances à partir d'un alignement des séquences nucléotidiques réalisé par le programme ClustalW.

Ainsi, compte tenu du fait que *S. avermitilis* est une espèce phylogénétiquement plus éloignée de *S. coelicolor* que ne l'est *S. ambofaciens*, la duplication des gènes *has* semble être un événement récent au cours de l'évolution, s'étant produit ultérieurement à la divergence entre les deux espèces *S. coelicolor* et *S. ambofaciens* et est conservée chez les trois isolats de *S. ambofaciens*.

- CHAPITRE B-III -

ANALYSE TRANSCRIPTIONNELLE DES  
GENES DUPLIQUES *HASR* ET *HASL*

## B-III.1 $\otimes$ ANALYSE QUALITATIVE DE LA TRANSCRIPTION DES GÈNES *HAS* PAR RT-PCR

Une première approche d'amplification par PCR des produits de transcription inverse (ou RT pour rerverse transcription) d'ARN totaux a été développée afin d'étudier l'expression des gènes *has* au niveau transcriptionnel. L'objectif était de développer un test qualitatif de l'expression des gènes *has* afin de déterminer les conditions d'expression de ces gènes.

### B-III.1.1 - EXPRESSION DES GÈNES *HASR* ET *HASL* CHEZ LA SOUCHE SAUVAGE *S. AMBOFACIENS* DSM40697

Après avoir réalisé une courbe de croissance de la souche sauvage *S. ambofaciens* DSM40697 (voir Matériel et Méthodes), les temps de culture retenus pour les points correspondant aux phases de croissance exponentielle, de transition (entrée en phase stationnaire) et stationnaire sont respectivement 9 h, 16 h et 24 h. Le choix de ces temps a été validé en réalisant la courbe de croissance à trois reprises.

Les ARN totaux ont été extraits aux trois temps de la courbe de croissance de la souche sauvage *S. ambofaciens* DSM40697. Des expériences de RT-PCR sur ces ARN ont alors été réalisées.

Dans un premier temps, l'expression globale des ORF *has*, c'est-à-dire simultanément des copies *hasL* et *hasR*, a été détectée par des expériences de PCR multiplex à l'aide des couples d'amorces A/B (localisé à l'intérieur de la région à 100% identique entre les deux loci *has*) et A1/A2, permettant d'amplifier le gène contrôle *recA*.

Les résultats obtenus permettent de conclure à l'expression de l'un ou des deux gènes *has* et ceci dans les trois phases de croissance (Figure R-35).

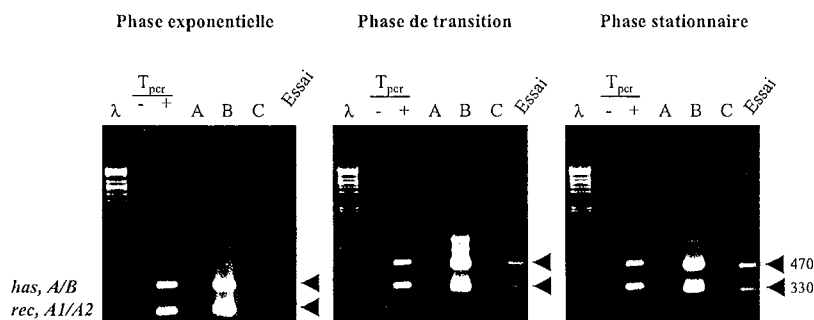
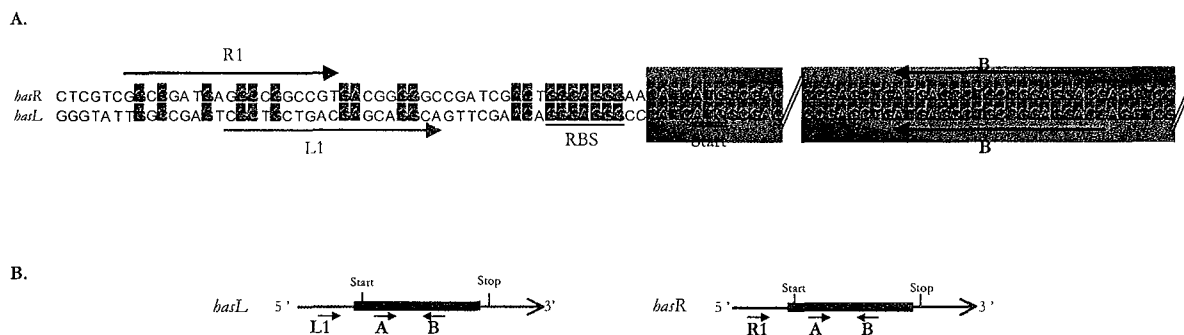


Figure R-35 : Mise en évidence de la transcription des gènes *recA* et *has* par RT-PCR chez la souche sauvage *S. ambofaciens* DSM40697 aux trois points de la courbe de croissance.  $T_{per}$  indique les témoins négatifs (-) et positifs (+) de PCR. Les témoins A, B et C ainsi que l'essai sont décrits dans le Matériel et Méthodes. Les flèches indiquent les bandes correspondant aux amplifiats spécifiques des loci *recA* (330 pb, amorces A1/A2) et *has* (470 pb, amorces A/B). Les autres bandes visibles dans les témoins B correspondent à des amplifications aspécifiques vraisemblablement dues à l'excès de matrice d'ADN dans l'échantillon d'ARN.

Dans un second temps, la divergence des séquences nucléotidiques caractérisée en amont (région 5') des ORF *has* a été mise à profit pour déceler l'expression spécifique de chacune des deux copies *hasR* et *hasL*. L'amorce choisie en amont du codon d'initiation de la traduction, à l'extérieur de la zone d'homologie, sera spécifique de la séquence à amplifier : il s'agit de R1 pour *hasR* et L1 pour *hasL* (Figure R-36). L'amorce B, utilisée en couple avec L1 et R1, appartient à la région identique à 100% en nucléotides. Ainsi, l'utilisation du couple d'amorces R1/B permettra de distinguer la copie *hasR* de la copie *hasL*, spécifiquement amplifiée avec le couple d'amorces L1/B en PCR.



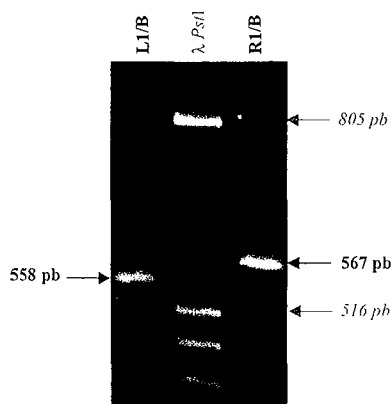
**Figure R-36 : Choix des amorces spécifiques des loci *hasR* et *hasL*.**

**A.** : Alignement des séquences des régions 5' des loci *hasR* et *hasL*.

**B.** Localisation des amorces de PCR aux loci *hasL* et *hasR*.

Start : codon d'initiation de la traduction, souligné d'un trait double. Stop : codon d'arrêt de la traduction. RBS (GGAGG) : Ribosome Binding Site, souligné d'un trait simple. Les nucléotides dans les boîtes grisées sont communs aux deux séquences. Les rectangles bleus correspondent aux régions à 100% identiques entre les deux loci *has*. Les amorces utilisées en PCR et leur orientation (5' vers 3') sont indiquées par des flèches.

Une électrophorèse en gel d'agarose des amplifiats permet de vérifier l'amplification spécifique de chacune des 2 copies. La taille des produits est respectivement de 567 pb et 558 pb pour *hasR* et *hasL* (Figure R-37).



**Figure R-37 : Discrimination des amplifiats spécifiques des loci *hasL* et *hasR* par l'utilisation des couples d'amorces spécifiques (L1/B et R1/B).** Les conditions de migration sont indiquées dans le matériel et méthodes. Seules les tailles des bandes du marqueur de taille (*PstI*) encadrant les amplifiats sont indiquées

Ces expériences ont permis de conclure à la transcription des deux gènes *hasR* et *hasL* dans les trois phases de croissance (Figure R-38).

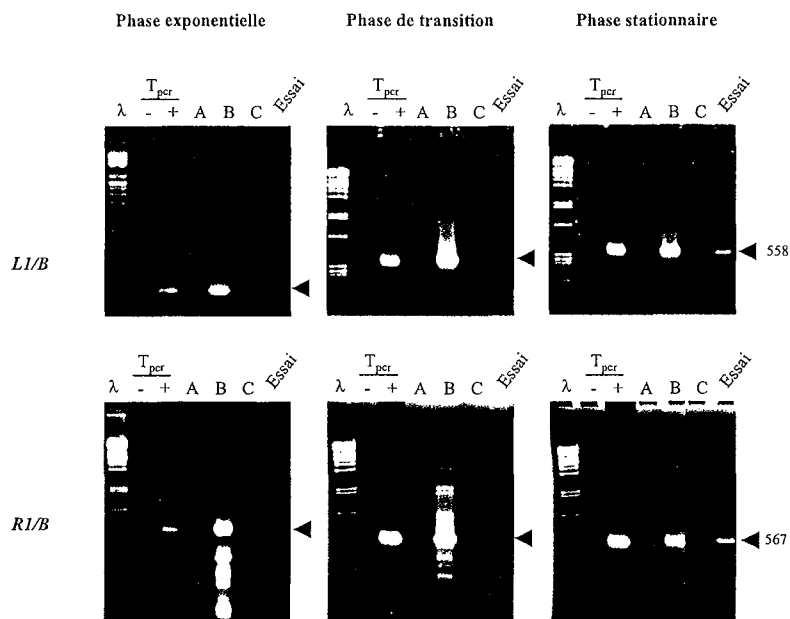


Figure R-38 : Mise en évidence de la transcription des gènes *hasL* et *hasR* par RT-PCR chez la souche sauvage *S. ambifaciens* DSM40697 aux trois points de la courbe de croissance. Les flèches indiquent les bandes correspondant aux amplifiats spécifiques des loci *hasL* (558 pb, couple d'amorces L1/B) et *hasR* (567 pb, couple d'amorces R1/B). Le reste du symbolisme est identique à celui de la Figure R-35.

### B-III.1.2 - EXPRESSION DES GENES *HASR* ET *HASL* ET DES CHIMERES *HAS855* ET *HAS135* AU COURS DE LA CROISSANCE

De même que pour la souche sauvage, la détermination des différents temps de culture, correspondant aux 3 phases de croissance étudiées, a été effectuée par la réalisation de courbes de croissance des souches mutantes NSAH135 et NSA855 (voir Matériel et Méthodes). Les temps définis sont les mêmes que pour la souche sauvage, les courbes étant comparables, à savoir 9 h, 16 h et 24 h respectivement pour les points en phases de croissance exponentielle, de transition et stationnaire.

Les ARN totaux ont été extraits aux trois temps prédéfinis de la courbe de croissance des souches mutantes *S. ambifaciens* NSAH135 et NSA855 et ont subi les étapes de la transcription inverse puis de PCR

Une amplification a tout d'abord été observée à l'aide du couple d'amorces A/B, ne distinguant pas les copies sauvages des chimères, pour les deux souches mutantes et dans les 3 phases de croissance révélant la transcription d'au moins une des deux copies *has*, sauvage et/ou chimérique (exemple des résultats obtenus chez la souche NSAH135 en Figure R-39).

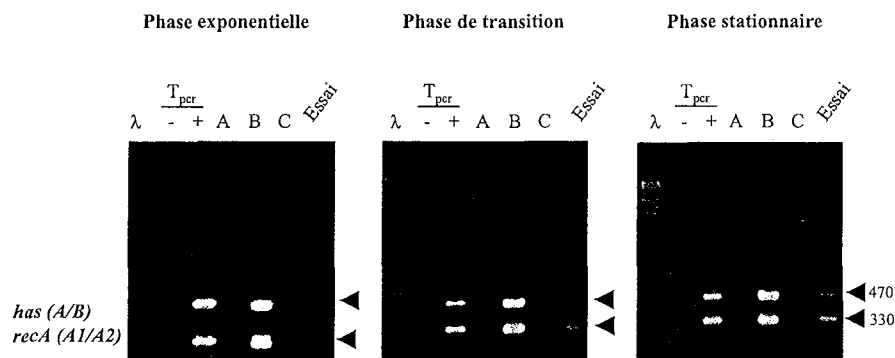


Figure R-39 : Mise en évidence de la transcription des gènes *recA* et *has* par RT-PCR chez la souche mutante *S. ambifaciens* NSAH135 aux trois points de la courbe de croissance. Les flèches indiquent les bandes correspondant aux amplifiats spécifiques des loci *recA* (330 pb, amorces A1/A2) et *has* (470 pb, amorces A/B). Le reste du symbolisme est identique à celui de la Figure R-35.

Les événements de recombinaison homologue, à l'origine des réarrangements chromosomiques chez les souches *S. ambifaciens* NSAH135 et NSA855, entraînent la formation de copies chimériques des gènes *has* (Figure R-25). Pour NSAH135, la chimère possède la région 5' du gène *hasL*. Ainsi, l'expression de *hasL* et de *has135* ne peut être distinguée avec les couples d'amorces R1/B et L1/B. La situation réciproque est générée pour NSA855. Cependant, les loci sauvage et chimérique divergent dans les régions 3' et notamment en aval du codon stop de traduction (Figures R-40), ce qui a permis de choisir des amorces spécifiques : il s'agit de R2 pour *hasR* et *has135* et L2 pour *hasL* et *has855*. Ces amorces seront utilisées en couple avec l'amorce D, choisie dans la région identique à 100% en nucléotides entre les différentes copies *has* et servant donc à l'amplification des copies sauvage et chimérique indistinctement.

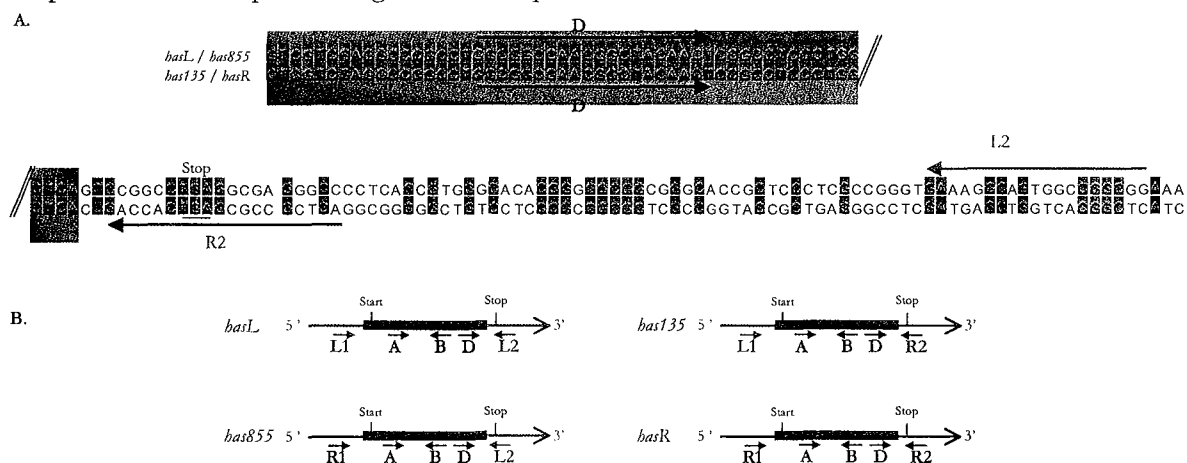


Figure R-40 : Choix des amorces spécifiques des loci *hasL* et *has135* chez la souche mutante NSAH135 et *hasR* et *has855* chez la souche mutante NSA855.

A. : Alignement des séquences des régions 3' des loci *hasL* ou *has855* et *hasR* ou *has135*.

B. Localisation des amorces de PCR aux loci *hasL*, *has135*, *has855* et *hasR*.

Start : codon d'initiation de la traduction. Stop : codon d'arrêt de la traduction, souligné d'un trait simple. Le reste du symbolisme est identique à celui de la Figure R-36.

Ainsi, chez la souche mutante NSAH135, deux réactions de PCR successives ont été réalisées : une première réaction utilisant le couple d'amorces D/L2 permettant d'amplifier spécifiquement le locus sauvage *hasL* et une seconde utilisant les amorces D/R2 permettant d'amplifier spécifiquement le locus chimérique *has135*.

Dans les deux expériences et dans les 3 phases de croissance, une amplification a été observée révélant la transcription du gène *hasL* et de la copie chimérique *has135* tout au long de la croissance de la souche mutante NSAH135 (Figure R-41).

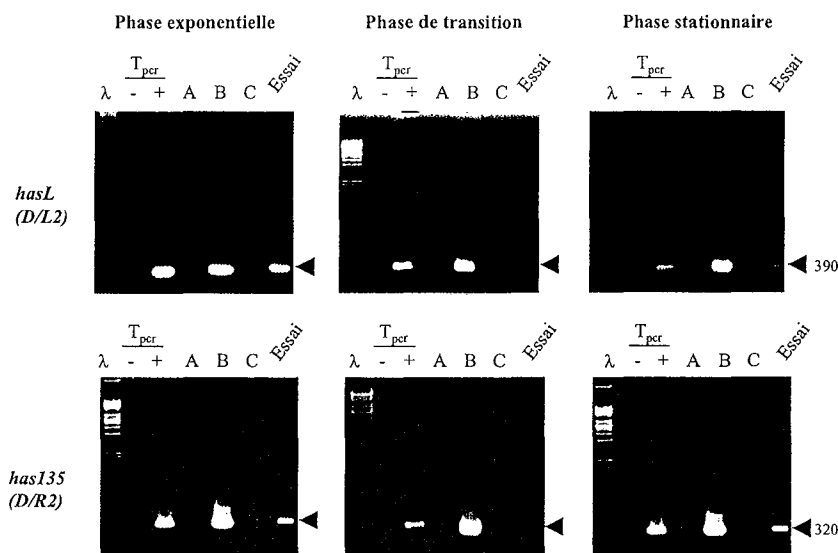


Figure R-41 : Mise en évidence de la transcription des gènes *hasL* et *has135* par RT-PCR chez la souche mutante *S. ambofaciens* NSAH135 aux trois points de la courbe de croissance. Les flèches indiquent les bandes correspondant aux amplifiats spécifiques des loci *hasL* (390 pb, couple d'amorces D/L2) et *has135* (320 pb, couple d'amorces D/R2). Le reste du symbolisme est identique à celui de la Figure R-35.

Des résultats analogues ont été obtenus pour la souche mutante NSA855, signifiant donc que la copie sauvage *hasR* de même que la copie chimérique *has855* s'expriment en phase exponentielle, en phase de transition et en phase stationnaire de croissance chez cette souche (Figure R-42).

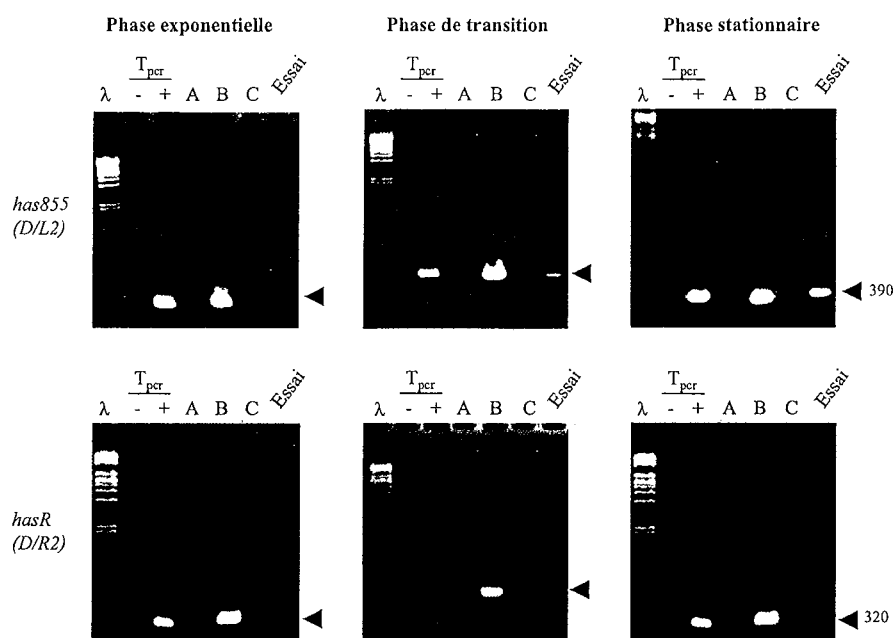


Figure R-42 : Mise en évidence de la transcription des gènes *hasR* et *has855* par RT-PCR chez la souche mutante *S. ambifaciens* NSA855 aux trois points de la courbe de croissance. Les flèches indiquent les bandes correspondant aux amplifiats spécifiques des loci *has855* (390 pb, couple d'amorces *D/L2*) et *hasR* (320 pb, couple d'amorces *D/R2*). Le reste du symbolisme est identique à celui de la Figure R-35.



## **B-III.2 $\sphericalangle$ IDENTIFICATION DES TRANSCRITS *HAS* ET ANALYSE DE LA REGULATION DE LEUR EXPRESSION**

Afin d'identifier les sites d'initiation de la transcription des gènes *has* (sites +1), des expériences d'extension d'amorce et de cartographie à la nucléase S1 ont été mises en oeuvre. Ces expériences ont tout d'abord été réalisées sur des ARN extraits à partir de cultures en phase stationnaire de croissance, puis la régulation de chaque transcrit a été analysée par ces mêmes approches à partir d'ARN extraits dans différentes phases de croissance ou suite à divers traitements.

### **B-III.2.1 – IDENTIFICATION DES SITES D'INITIATION DE LA TRANSCRIPTION ET DES PROMOTEURS DES GENES *HAS***

#### B-III.2.1.1 – IDENTIFICATION DES +1 DES GENES *HAS* PAR UNE APPROCHE D'EXTENSION D'AMORCES

Afin de ne présumer d'aucune localisation des promoteurs putatifs des gènes *has*, les oligonucléotides servant d'amorces à la reverse transcriptase devaient être choisis à l'intérieur des phases codantes des gènes *has*. La région d'homologie entre les deux gènes *has* débutant quelques nucléotides avant le codon d'initiation des ORF, les oligonucléotides choisis étaient communs aux deux séquences *hasR* et *hasL* et ne permettaient donc pas de distinguer les arrêts de la transcriptase inverse dus à un site d'initiation de la transcription de *hasR* plutôt que de *hasL* et inversement.

Cependant les souches mutantes transloquées NSAH135 et NSA855 nous ont servi d'outils dans cette analyse. En effet, les régions promotrices des gènes *has* étant homogénéisées suite au réarrangement présent chez ces souches, les sites d'initiation de la transcription d'un seul type de région promotrice ne peuvent être détectés à partir d'ARN de ces souches (Figure R-43 A.).

L'analyse des résultats des expériences d'extension d'amorce réalisées sur des ARN de la souche sauvage parallèlement à ceux obtenus sur des ARN de chacune de ces deux souches mutantes ont alors permis d'attribuer chaque arrêt à l'un ou l'autre des deux gènes *has*.

La Figure R-43 B présente une partie des résultats d'extension d'amorce obtenus à l'aide de l'oligonucléotide E, localisé à l'intérieur de la phase codante des gènes *has*, 49 pb en aval du codon d'initiation de la traduction.

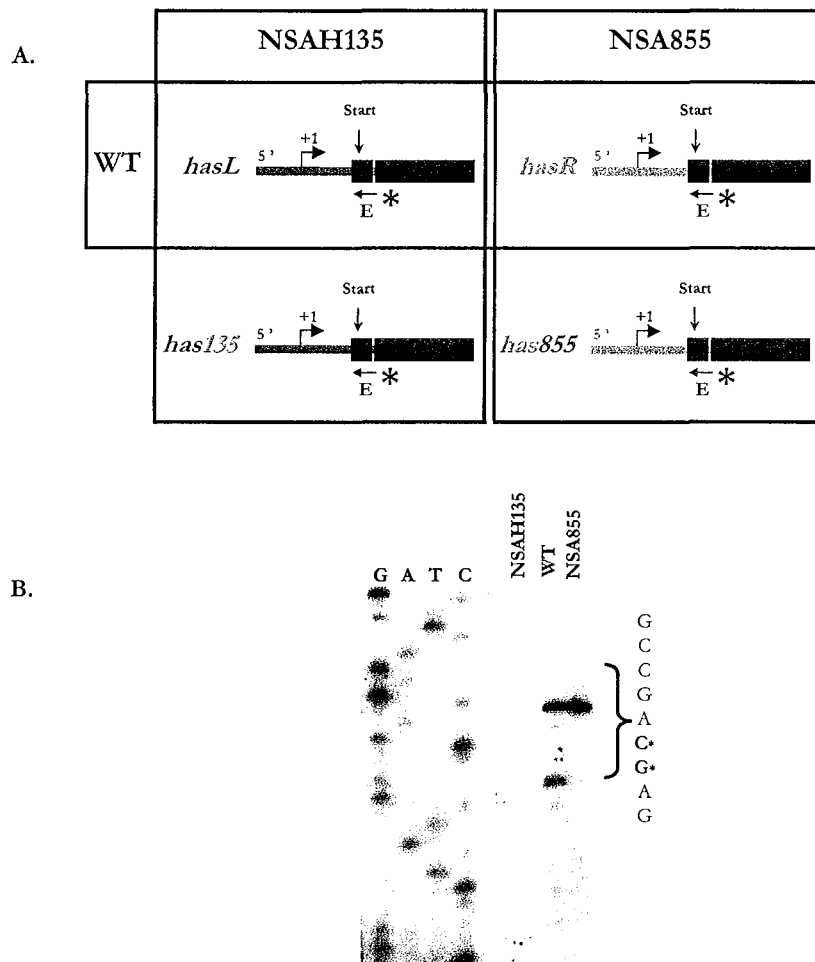


Figure R-43 : Identification des sites d'initiation de la transcription des gènes *has* par une approche d'extension d'amorces : exemple du transcrit initié au site *hasRp1*.

A. Illustration de l'homogénéisation des régions promotrices des gènes *has* chez les mutants NSAH135 et NSA855.

B. Résultat d'une expérience d'extension d'amorce, à l'aide de l'oligonucléotide E, sur des ARN des souches sauvages et mutantes NSAH135 et NSA855, extraits en phase stationnaire de croissance.

L'arrêt de la transcriptase inverse détecté sur des ARN de la souche sauvage est également retrouvé lorsque des ARN de la souche mutante NSA855, présentant des régions promotrices de type *hasR*, sont analysés mais n'existe pas dans le cas de la souche mutante NSAH135 où les deux copies *has* présentent des régions 5' de type *hasL*.

Ainsi, cette expérience permet d'identifier un site d'initiation de la transcription du gène *hasR*, localisé à 53 pb du codon start (*hasRp1*).

D'autres migrations de ces mêmes réactions d'extension d'amorces ainsi que de réactions réalisées à l'aide d'autres oligonucléotides plus internes nous ont permis de détecter plusieurs autres arrêts de la transcriptase inverse.

Le détail de la détermination des sites d'initiation de la transcription des gènes *has* et de l'analyse des séquences promotrices putatives ne sera présenté que dans le paragraphe relatant les résultats de cartographie à la nucléase S1 (paragraphe ci-après), sachant que ces sites +1 ont été retenus comme valides lorsqu'ils étaient retrouvés à l'aide d'au moins deux oligonucléotides différents lors des expériences d'extension d'amorces et confirmés par les expériences de cartographie à la nucléase S1.

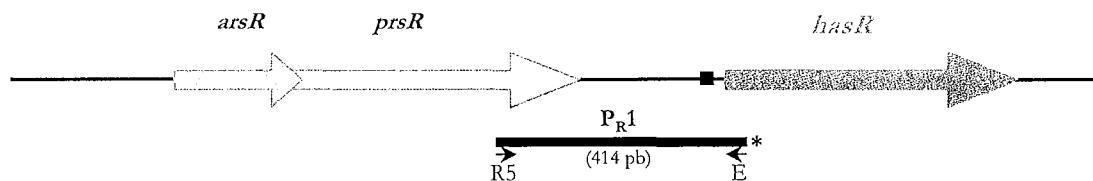
### B-III.2.1.2 – IDENTIFICATION DES SITES +1 DES GENES *HAS* PAR UNE APPROCHE DE CARTOGRAPHIE A LA NUCLEASE S1 ET ANALYSE DES SEQUENCES PROMOTRICES

Lors des expériences de cartographie à la nucléase S1, les différentes sondes utilisées ont été réalisées par PCR (voir Matériel et Méthodes).

#### B-III.2.1.2.1 - Les transcrits du gène *hasR*

##### B-III.2.1.2.1.1 – Le transcrit initié au site *hasRp1* et son promoteur de type $\sigma^B$

Pour le gène *hasR*, la première sonde, appelée P<sub>R</sub>1, correspond à une région de 414 pb chevauchant la région intergénique entre les gènes *hasR* et *prsR*. Cette sonde a été fabriquée par PCR à l'aide de l'oligonucléotide R5 localisé à l'intérieur du gène *prsR* et 364 pb en amont du codon d'initiation de la traduction du gène *hasR*, et de l'oligonucléotide E précédemment décrit (Figure R-44).

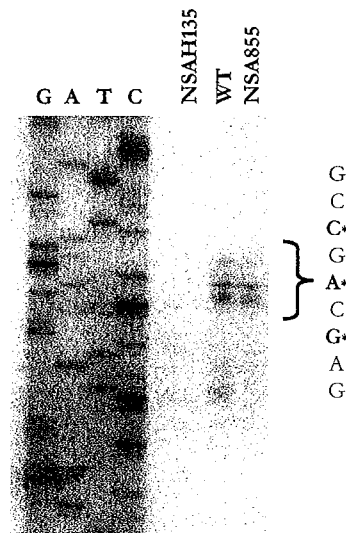


**Figure R-44 : Représentation schématique de la localisation de la sonde P<sub>R</sub>1 sur le locus *hasR*.**  
Le carré noir symbolise le site de fixation du ribosome (RBS). Les petites flèches noires représentent les oligonucléotides utilisés lors de la réaction de PCR donnant naissance à la sonde P<sub>R</sub>1. L'astérisque indique le marquage radioactif de la sonde, dû au marquage de l'oligonucléotide E.

Des expériences de cartographie à la nucléase S1 réalisées sur des ARN de la souche sauvage à l'aide de cette sonde P<sub>R</sub>1 donnent naissance à deux fragments protégés : un fragment de grande taille compatible avec une protection de la sonde et un fragment de 102 nucléotides (non montré).

Le fragment de 102 nucléotides est le résultat de l'hybridation d'un transcrit initié 53 pb en amont du codon d'initiation de la traduction du gène *hasR* sur la sonde P<sub>R</sub>1 (puisque l'oligonucléotide E est localisé 49 pb en aval de ce même codon : 102-49=53). Ce premier site d'initiation de la transcription du gène *hasR*, également détecté par les expériences d'extension d'amorces (voir paragraphe B-III.2.1.1) a été nommé *hasRp1*.

De la même façon que pour les expériences d'extension d'amorces, ce départ de transcription spécifique du gène *hasR* est détecté chez la souche NSA855 (homogénéisée pour des régions promotrices de type *hasR*) alors qu'il ne l'est pas chez la souche NSAH135 (homogénéisée pour des régions promotrices de type *hasL*, Figure R-45).



**Figure R-45 : Identification des sites d'initiation de la transcription des gènes *has* par une approche de cartographie à la nucléase S1 : exemple du transcrit initié au site *hasRp1*.**

Le film montré présente le résultat d'une expérience de protection à la nucléase S1 réalisée sur des ARN des souches sauvages et mutantes NSAH135 et NSA855, extraits en phase stationnaire de croissance.

L'analyse des séquences promotrices putatives en amont de ce site d'initiation de la transcription *hasRp1* a ensuite été entreprise. Une séquence similaire au consensus des promoteurs reconnus par le facteur de réponse au stress général  $\sigma^B$  de *B. subtilis* a ainsi pu être identifiée, dont le motif -10 est localisé à 11 pb du site +1 et l'espacement entre les deux motifs -10 et -35 est de 13 pb (Figure R-46 A.). Un promoteur du même type a été identifié en amont du gène *sigB*, homologue du gène *hasR*, chez *S. coelicolor* (Cho *et al.*, 2001).

Un alignement de ce promoteur putatif avec le consensus défini chez *B. subtilis* est présenté en Figure R-46 B.

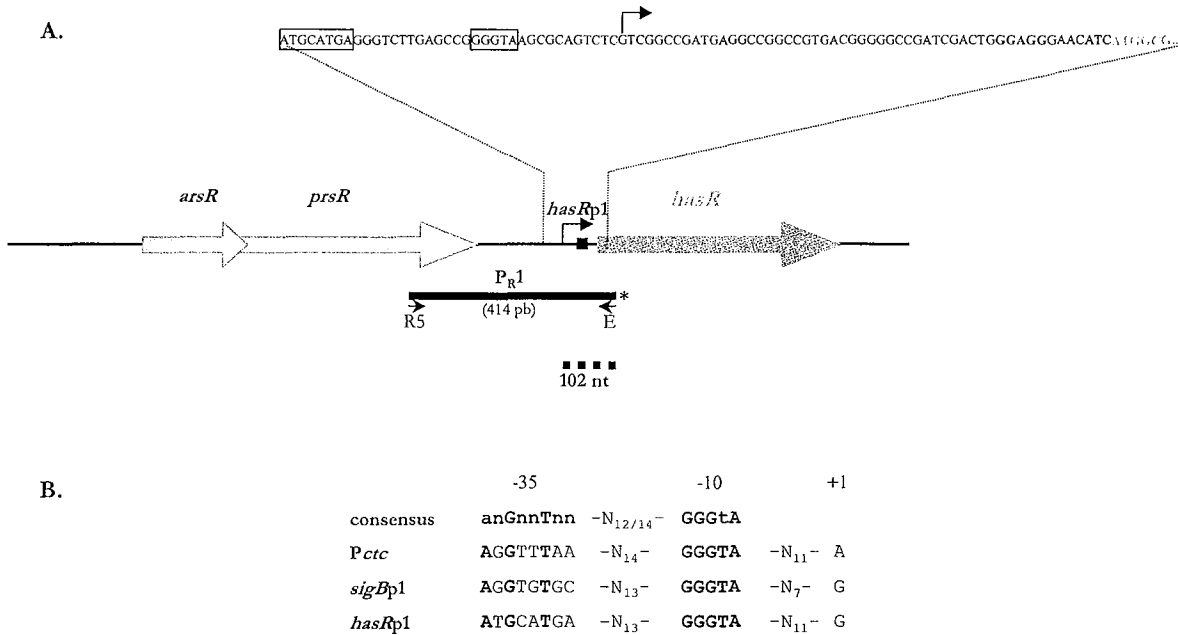


Figure R-46 : Caractérisation du transcrit initié au site *hasRp1* et de son promoteur de type  $\sigma^B$ .

A. Représentation schématique des caractéristiques du transcrit initié au site *hasRp1*. Le trait en pointillés indique le fragment protégé détecté par les expériences de cartographie à la nucléase S1. La séquence promotrice du gène *hasR* est indiquée. La flèche symbolise le site d'initiation de la transcription *hasRp1* et les rectangles noirs encadrent les séquences -35 et -10 du promoteur. Les nucléotides indiqués en rouge correspondent à la séquence codante du gène *hasR*. Le rectangle hachuré symbolise le début de la séquence homologue entre les gènes *has*.

B. Alignement de différents promoteurs similaires à celui identifié en amont de *hasRp1*.

B-III.2.1.2.1.2 – Le transcrit initié au site *hasRp2* et son promoteur de type  $\sigma^B$

Le fragment de grande taille obtenu lors de cette même expérience de cartographie à la nucléase S1 à l'aide de la sonde P<sub>R1</sub> résulterait de la protection de la totalité de la sonde par un ARN probablement polycistronique (non montré). Compte tenu du fait que les gènes *arsR* et *prsR* sont chevauchants, le site +1 correspondant à ce transcrit a été recherché en amont du gène *arsR*.

Ainsi, la sonde P<sub>R2</sub> a été fabriquée par PCR à l'aide du couple d'oligonucléotides *arsR1/arsR2* localisés à 327 et 43 pb respectivement en amont et en aval du codon d'initiation de la traduction du gène *arsR* (Figure R-47).

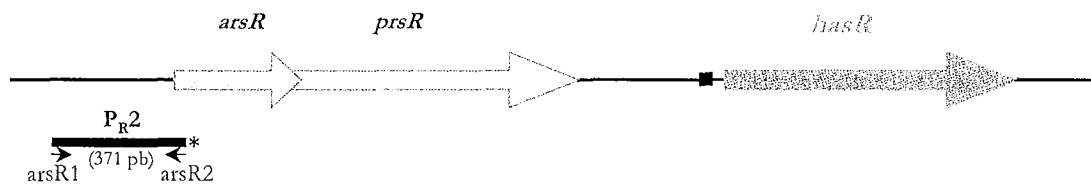


Figure R-47 : Représentation schématique de la localisation de la sonde P<sub>R2</sub> sur le locus *hasR*. Le symbolisme utilisé est identique à celui de la Figure R-44.

Cette sonde génère un fragment protégé unique de 157 nucléotides démontrant la présence d'un site d'initiation de la transcription, nommé *hasRp2*, localisé 114 pb en amont du codon d'initiation de la traduction du gène *arsR* (non montré, voir paragraphe B-III.2.2).

En amont de ce site d'initiation de la transcription, une séquence similaire au consensus des promoteurs reconnus par le facteur de réponse au stress général  $\sigma^B$  de *B. subtilis* a également pu être identifiée (Figure R-48 A.), mais il est important de noter la moindre conservation de la boîte -10 (4 nucléotides sur 5 conservés) et l'espacement entre les deux motifs -10 et -35 légèrement atypique (une base plus court que le consensus, voir l'alignement en Figure R-48 B.).

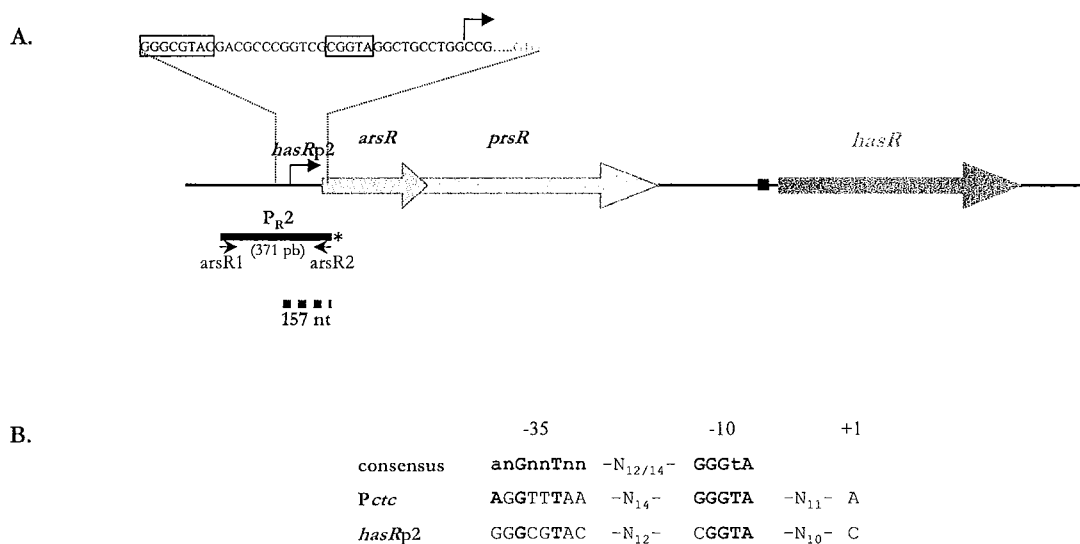


Figure R-48 : Caractérisation du transcrit initié au site *hasRp2* et de son promoteur de type  $\sigma^B$ .

A. Représentation schématique des caractéristiques du transcrit initié au site *hasRp2*. Les nucléotides indiqués en rouge correspondent à la séquence codante du gène *arsR*. Le reste du symbolisme est identique à celui de la Figure R-46.

B. Alignement de différents promoteurs similaires à celui identifié en amont de *hasRp2*.

### B-III.2.1.2.1.3 – Les promoteurs putatifs du gène *hasR* reconnus par *WhiG*

L'analyse des séquences promotrices putatives dans la région 5' du gène *hasR* a également révélé la présence de deux promoteurs putatifs reconnus par le facteur  $\sigma$  *WhiG*, localisés cependant en aval des sites d'initiation de la transcription détectés (Figure R-49 A. et alignement en Figure R-49 B.)

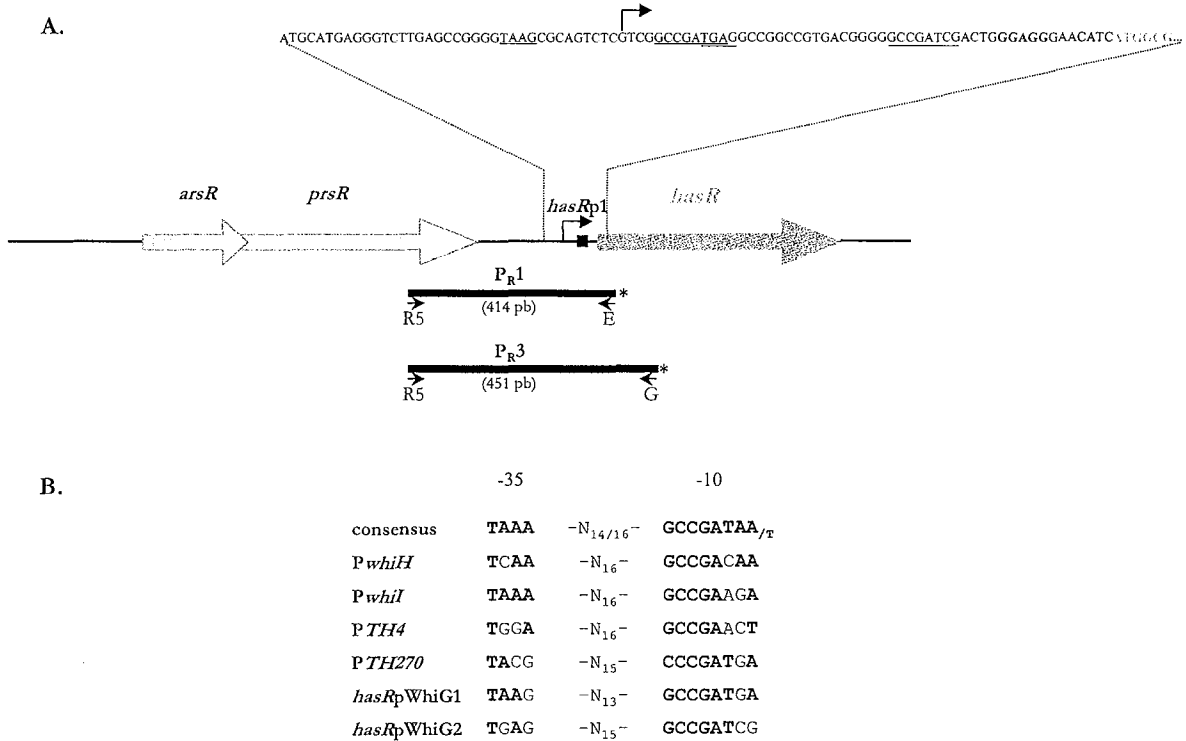


Figure R-49 : Les promoteurs putatifs reconnus par le facteur  $\sigma$  WhiG en amont du gène *hasR*.

A. Représentation schématique de la localisation des sondes P<sub>R1</sub> et P<sub>R3</sub> sur le locus *hasR*. Les boîtes -35 et -10 des promoteurs de type  $\sigma$  WhiG sont soulignées. Le reste du symbolisme utilisé est identique à celui des Figures R-44 et R-46.

B. Alignement de différents promoteurs, putatifs ou non, de type  $\sigma$  WhiG.

Afin de confirmer que ces promoteurs, ne donnant pas naissance à un transcrit détectable à l'aide la sonde P<sub>R1</sub>, n'étaient effectivement pas fonctionnels, une sonde, nommée P<sub>R3</sub>, a alors été fabriquée par PCR à l'aide du même oligonucléotide R5 en 5' et d'un oligonucléotide en 3' appelé G, localisé plus à l'intérieur du gène *hasR* (96 pb en aval du codon start, contrairement aux 49 pb pour l'oligonucléotide E, Figure R-49 A.). Ainsi, cette sonde générerait des fragments protégés de 93 et 118 pb, détectables lors de nos expériences, si un transcrit était initié via ces promoteurs de type  $\sigma^{\text{WhiG}}$ .

Le seul fragment observé lors d'expériences de cartographie à la nucléase S1 utilisant cette sonde P<sub>R3</sub> sur des ARN extraits en différentes phases de croissance en milieu HT liquide ou solide correspond au transcrit initié au site *hasRp1* (non montré). Cependant, nous ne pouvons pas affirmer que ces promoteurs ne soient pas actifs dans d'autres conditions que celles que nous avons testées.

De façon intéressante, l'analyse de la séquence en amont du promoteur *sigBp1* du gène *sigB* chez *S. coelicolor* a révélé la conservation d'un des deux promoteurs putatifs reconnus par le facteur  $\sigma$  WhiG. De plus, le chevauchement entre ce promoteur putatif et le promoteur *sigBp1* est également retrouvé (Figure R-50).

*S. a hasR* ATGATCAGGGTCTTGAGCCGAGGTAAGCGCAGTCTCGTCCCCGATGAGCCGGCCGTGACGGGGCCGATCGACTGGGAGGGAACATC...  
*S. c sigB* ACCTGACCGGTGCTTGACCCGAGGTAAGCGGATCGGTCTCGCCGACAAGGTGACAACGACGACCGATCCACCGGAGGGCACATC...

Figure R-50 : Alignement des régions promotrices des gènes *hasR* de *S. ambifaciens* et *sigB* de *S. coelicolor*

Les boîtes violettes encadrent les séquences -35 et -10 des promoteurs *hasRp1* et *sigBp1*. L'astérisque symbolise le site d'initiation de la transcription associé à ces promoteurs. Les promoteurs putatifs reconnus par le facteur  $\sigma$  WhiG sont soulignés. Le RBS est indiqué en gras (GGAGG). Les nucléotides indiqués en rouge correspondent à la séquence codante des deux gènes.

### B-III.2.1.2.2 - Les transcrits du gène *hasL*

Pour le gène *hasL*, l'unique sonde utilisée de 446 pb, appelée  $P_L$ , correspond à la région en amont du gène *hasL*. Cette sonde a été fabriquée par PCR à l'aide de l'oligonucléotide L5 localisé 396 pb en amont du codon d'initiation de la traduction du gène *hasL*, et de l'oligonucléotide E précédemment décrit (Figure R-51).

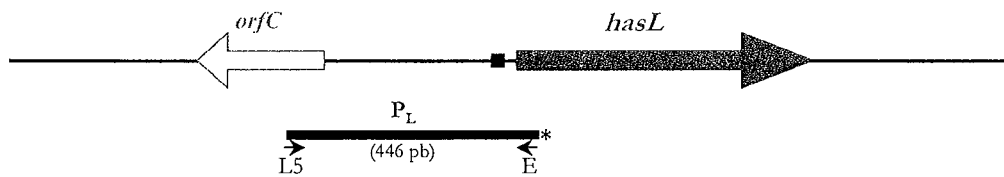


Figure R-51 : Représentation schématique de la localisation de la sonde  $P_L$  sur le locus *hasL*.

Le carré noir symbolise le site de fixation du ribosome (RBS). Les petites flèches noires représentent les oligonucléotides utilisés lors de la réaction de PCR donnant naissance à la sonde  $P_L$ . L'astérisque indique le marquage radioactif de la sonde, dû au marquage de l'oligonucléotide E.

Les expériences de cartographie à la nucléase S1 réalisées sur des ARN de la souche sauvage à l'aide de cette sonde  $P_L$  donnent naissance à deux fragments protégés de 92 et 136 nucléotides (non montré, voir paragraphe B-III.2.2).

Ces deux fragments protégés démontrent l'existence de deux transcrits du gène *hasL* (Figure R-52) :



B-III.2 Identification des transcrits *has* et analyse de la régulation de leur expression

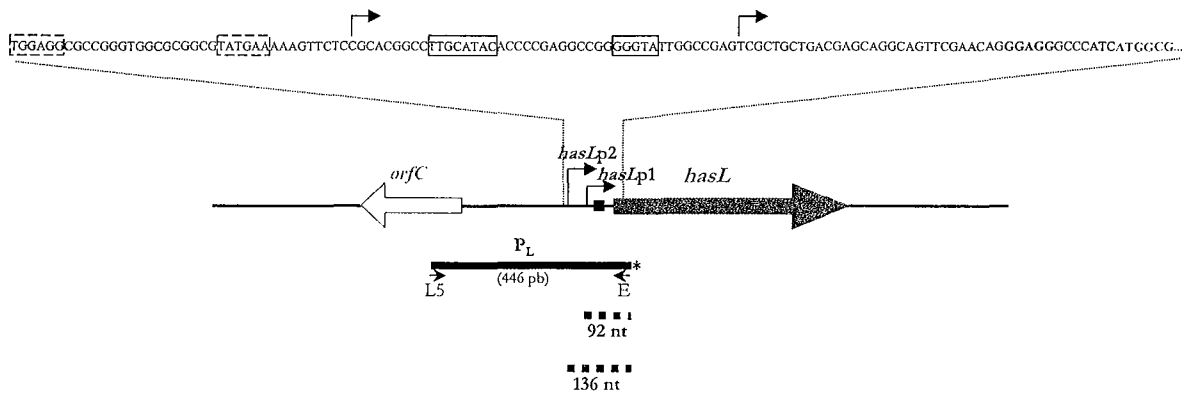


Figure R-52 : Caractérisation des transcrits initiés aux sites *hasLp1* et *hasLp2*.

Les traits en pointillés indiquent les fragments protégés détectés par les expériences de cartographie à la nucléase S1. La séquence promotrice du gène *hasL* est indiquée. Les flèches symbolisent les sites d'initiation de la transcription *hasLp1* et *hasLp2* et les rectangles noirs (traits pleins et traits pointillés respectivement) encadrent les séquences -35 et -10 des promoteurs. Les nucléotides indiqués en bleu correspondent à la séquence codante du gène *hasL*.

B-III.2.1.2.2.1 – Le transcrit initié au site *hasLp1* et son promoteur de type  $\sigma^B$

Le fragment de 92 nucléotides correspond à un transcrit initié au site *hasLp1*, localisé 43 pb en amont du codon d'initiation de la traduction du gène *hasL*.

L'analyse des séquences promotrices putatives en amont de ce site d'initiation de la transcription a ensuite également été entreprise. Comme pour *hasRp1*, une séquence similaire au consensus des promoteurs reconnus par le facteur de réponse au stress général  $\sigma^B$  de *B. subtilis* a pu être identifiée en amont de *hasLp1* (Figure R-53 A. et alignement en Figure R-53 B.).

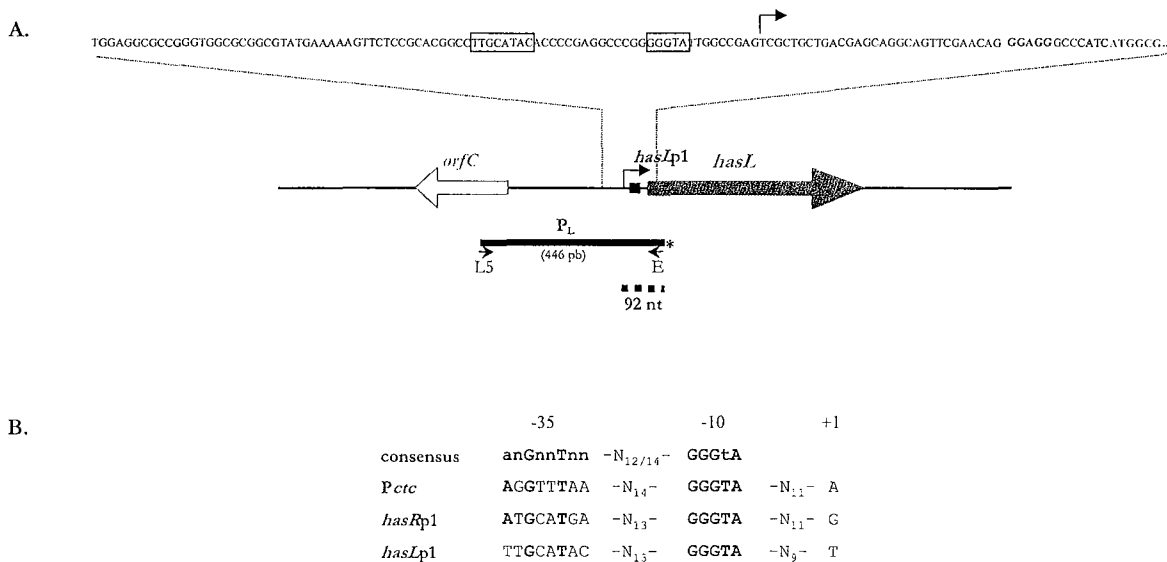


Figure R-53 : Caractérisation du transcrit initié au site *hasLp1* et de son promoteur de type  $\sigma^B$ .

A. Représentation schématique des caractéristiques du transcrit initié au site *hasLp1*. Le symbolisme est identique à celui de la Figure R-52.

B. Alignement de différents promoteurs similaires à celui identifié en amont de *hasLp1*.

B-III.2.1.2.2.2 – Le transcrit initié au site *hasLp2* et son promoteur de type  $\sigma^{\text{HrdB}}$

Le fragment protégé de 136 nucléotides, révélé à l'aide de la sonde P<sub>L</sub>, démontre l'existence d'un second site d'initiation de la transcription du gène *hasL*, *hasLp2*, localisé 87 pb en amont du codon d'initiation de la traduction (Figure R-54 A.).

En amont du site *hasLp2*, une séquence homologe aux promoteurs reconnus par le facteur  $\sigma$  général chez *S. coelicolor*,  $\sigma^{\text{HrdB}}$ , a pu être identifiée (alignement en Figure R-54 B.).

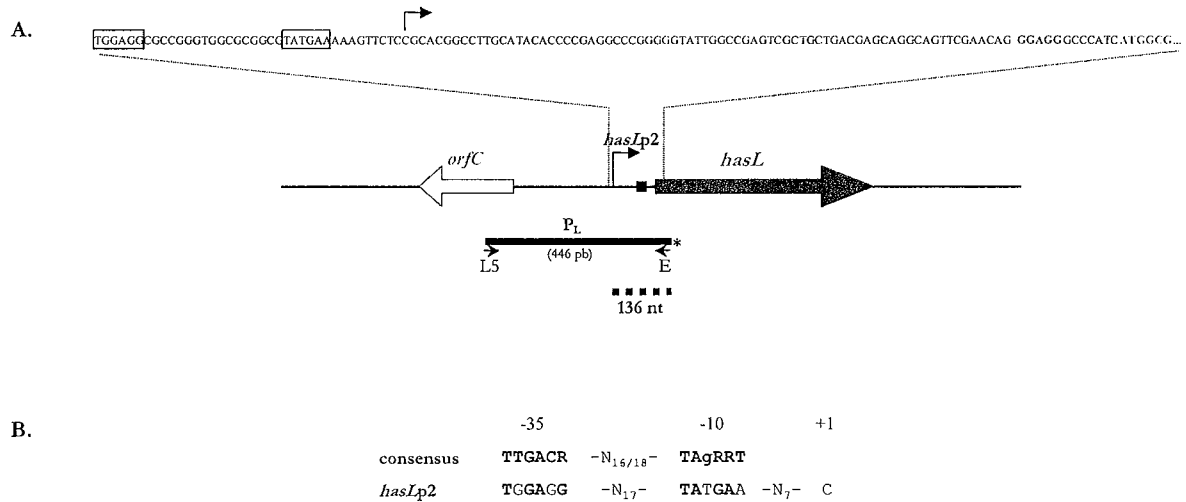


Figure R-56 : Caractérisation du transcrit initié au site *hasLp2* et de son promoteur de type  $\sigma^{\text{HrdB}}$ .

A. Représentation schématique des caractéristiques du transcrit initié au site *hasLp2*. Le symbolisme est identique à celui de la Figure R-52.

B. Alignement du consensus des promoteurs de type  $\sigma^{\text{HrdB}}$  avec le promoteur identifié en amont de *hasLp2*.

## B-III.2.2 - REGULATION DE L'EXPRESSION DES TRANSCRITS *HAS* AU COURS DE LA CROISSANCE

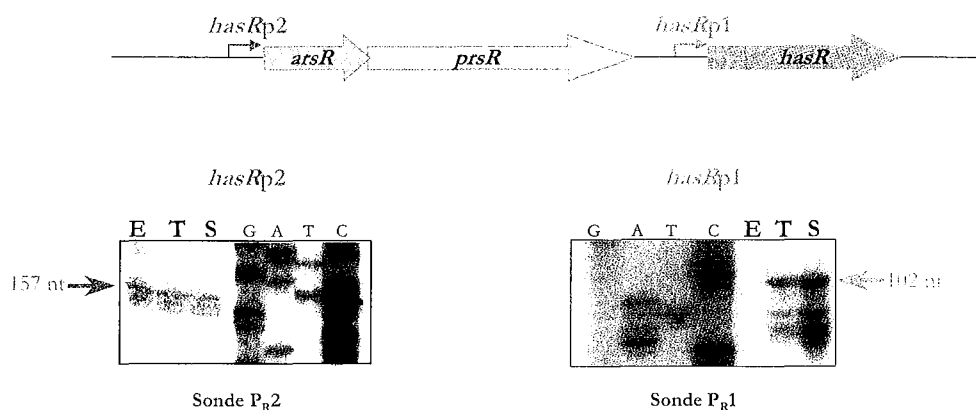
### B-III.2.2.1 - IDENTIFICATION DE LA REGULATION DE LA TRANSCRIPTION DES GENES *HAS* AU COURS DE LA CROISSANCE PAR CARTOGRAPHIE A LA NUCLEASE S1

L'expression de chaque transcrit *has* a ensuite été analysée au cours de la croissance en milieu HT liquide.

Des ARN extraits à partir de culture en phase exponentielle (après 9 heures de culture), de transition (après 16 heures de culture) et stationnaire (après 24 heures de culture) de croissance ont ainsi été analysés par des expériences de cartographie à la nucléase S1 à l'aide des mêmes sondes que précédemment.

En ce qui concerne le gène *hasR*, il est apparu que les deux transcrits initiés aux sites *hasRp1* et *hasRp2* présentaient un profil d'expression différent au cours de la croissance (Figure R-55) avec :

- une induction de la transcription à partir du site *hasRp1* au cours de la croissance ;
- et une expression constante du transcrit initié au site *hasRp2*.

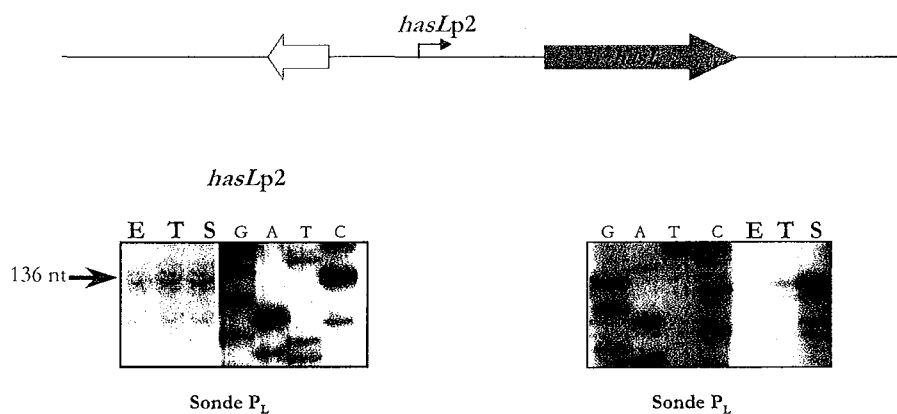


**Figure R-55 : Régulation des transcrits du gène *hasR* au cours de la croissance.**

Cette figure présente les résultats des expériences de cartographie à la nucléase S1 réalisées à l'aide des sondes P<sub>R2</sub> et P<sub>R1</sub> sur des ARN extraits dans les trois phases de croissance (E, exponentielle ; T, transition ; S, stationnaire) d'une culture en milieu HT liquide.

De la même manière, ces expériences ont permis de montrer pour le gène *hasL* que l'expression du transcrit initié au site *hasLp1* était induite au cours de la croissance alors que celle du transcrit initié au site *hasLp2* était relativement constante (Figure R-56).

Il est cependant important de noter que l'exposition des films autoradiographiques a été plus longue dans le cas des expériences réalisées avec la sonde  $P_L$  qu'avec la sonde  $P_R1$ , suggérant que le gène *hasR* serait transcrit à un niveau plus élevé que le gène *hasL*.



**Figure R-56 : Régulation des transcrits du gène *hasL* au cours de la croissance.**

Cette figure présente les résultats des expériences de cartographie à la nucléase S1 réalisées à l'aide de la sonde  $P_L$  sur des ARN extraits dans les trois phases de croissance (E, exponentielle ; T, transition ; S, stationnaire) d'une culture en milieu HT liquide.

### B-III.2.2.2 – QUANTIFICATION DES NIVEAUX D'INDUCTION DE LA TRANSCRIPTION DES GENES *HAS* AU COURS DE LA CROISSANCE PAR PCR QUANTITATIVE

Afin de confirmer ces résultats et dans le but de quantifier le niveau d'expression de chaque transcrit, une approche de PCR quantitative en temps réel sur des ADNc produits par reverse transcription des mêmes ARN extraits au cours des différentes phases de croissance a été développée (voir Matériel et Méthodes).

#### B-III.2.2.2.1 – Principes

##### B-III.2.2.2.1.1 – Principe de la quantification de chaque transcrit *has*

Les couples utilisés dans ces expériences de PCR quantitative sont nommés (Figure R-57) :

- Rqu1/Rqu2 spécifique des transcrits initiés à *hasRp1* et *hasRp2*
- Rqu3/Rqu4 spécifique du transcrit initié à *hasRp2*
- Lqu1/Lqu2 spécifique des transcrits initiés à *hasLp1* et *hasLp2*
- Lqu3/Lqu2 spécifique du transcrit initié à *hasLp2*

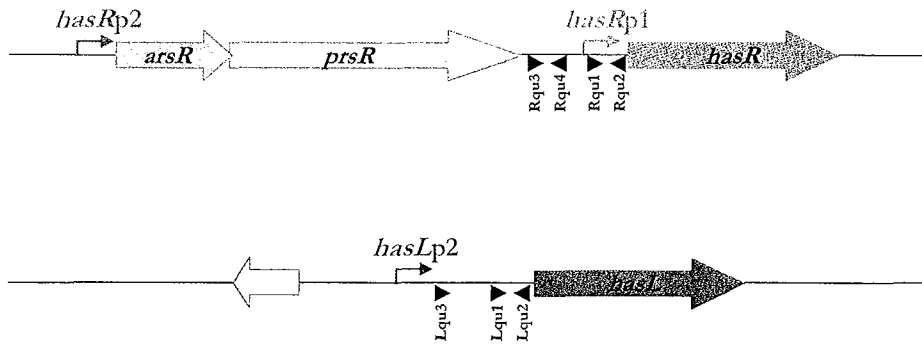


Figure R-57 : Localisation des différentes amorces utilisées dans les réactions de RT-PCR quantitative par rapport aux sites d'initiation de la transcription des gènes *has*.

L'expression des transcrits initiés spécifiquement au niveau des promoteurs *hasRp1* et *hasLp1* a été quantifiée par différence entre le niveau global d'expression du gène *has* considéré (obtenu à l'aide du couple d'amorce Rqu1/Rqu2 pour *hasR* et Lqu1/Lqu2 pour *hasL*) soustrait du niveau d'expression spécifique du transcrit *hasRp2* et *hasLp2* respectivement (obtenu à l'aide du couple d'amorce Rqu3/Rqu4 pour *hasR* et Lqu3/Lqu2 pour *hasL*). Un exemple d'un tel calcul est présenté en Figure R-58.

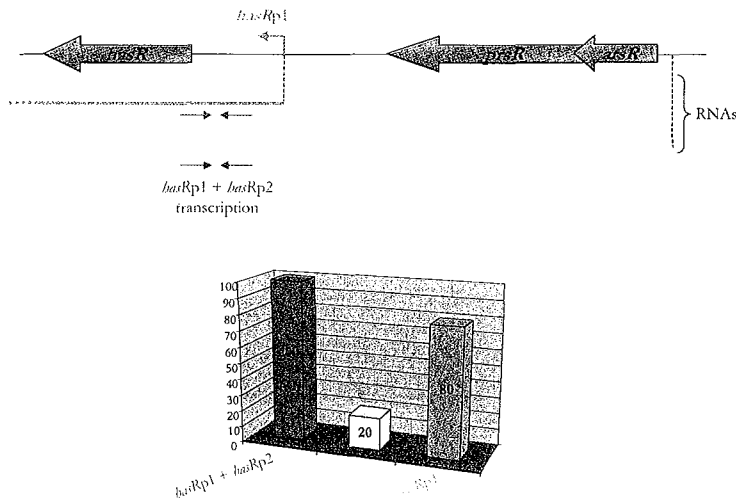


Figure R-58 : Principe de la quantification spécifique de chaque transcrit *has*.

Cette figure présente l'exemple d'un calcul fictif permettant de déduire le niveau d'expression du transcrit *hasRp1*.

Dans ces expériences, l'homologue chez *S. ambofaciens* du gène *brdB*, codant le facteur  $\sigma$  principal chez *S. coelicolor*, a été utilisé comme contrôle interne puisque présente la caractéristique d'être transcrit à un niveau relativement constant au cours de la croissance.

L'expression de chaque transcrit *has* a donc été déterminée de façon relative à celle du transcrit *brdB*, en utilisant la phase exponentielle de croissance comme référence.

B-III.2.2.2.1.2 – Traitement mathématique des données

La technologie de PCR quantitative passe par un traitement mathématique des données et permet ainsi une certaine liberté au niveau de la forme des résultats obtenus (voir Matériel et Méthodes). Ainsi, deux types de calcul en particulier ont été réalisés :

□ Le premier prend en compte le niveau d'expression de chaque transcrit *has* indépendamment et ne permet donc pas de comparer les niveaux de transcription des deux gènes *has* (Tableau R-8). Dans ce traitement des données, la valeur correspondant au niveau d'expression de chaque transcrit en phase exponentielle est arbitrairement fixée à 100.

	Exponentielle	Transition	Stationnaire
<i>hasIp2</i>	100	90	128
<i>hasIp1</i>	100	390	1100
<i>hasRp2</i>	100	119	229
<i>hasRp1</i>	100	3000	10900

Tableau R-8 : Evolution du niveau d'expression de chaque transcrit *has*.

□ Le second type de calcul donne l'abondance relative de chacun des 4 transcrits *has* dans chaque phase de croissance (Tableau R-9). Ainsi, la valeur de la transcription globale des deux gènes *has*, via leurs deux promoteurs respectifs, est arbitrairement fixée à 100 dans cette analyse.

	Exponentielle	Transition	Stationnaire
<i>haspI2</i>	1	0,9	1,3
<i>hasIp1</i>	9	35	99
<i>hasRp2</i>	2	2,4	4,6
<i>hasRp1</i>	88	2640	9592

Tableau R-9 : Abondance relative de chaque transcrit *has*.

Cette analyse par PCR quantitative a ainsi permis de confirmer les régulations de chaque transcrit *has* au cours de la croissance préalablement montrées par les expériences de cartographie à la nucléase S1.

**B-III.2.2.2.2 – Quantification de la transcription du gène *hasR****B-III.2.2.2.2.1 – Evolution de la quantité de chaque transcrit du gène *hasR* au cours de la croissance*

En effet, il est apparu, concernant le gène *hasR*, que la transcription initiée au site *hasRp2* était relativement constante (variation maximale d'un facteur 2), alors que celle initiée au site *hasRp1* était induite au cours de la croissance. De plus, cette induction a pu être quantifiée et estimée d'un facteur 30 lors du passage de la phase exponentielle à la phase de transition. L'induction la plus remarquable de la transcription du gène *hasR* est observée en phase stationnaire de croissance, puisque le niveau d'expression du transcrit initié au site *hasRp1* est 110 fois plus élevé qu'en phase exponentielle (Figure R-59 B.).

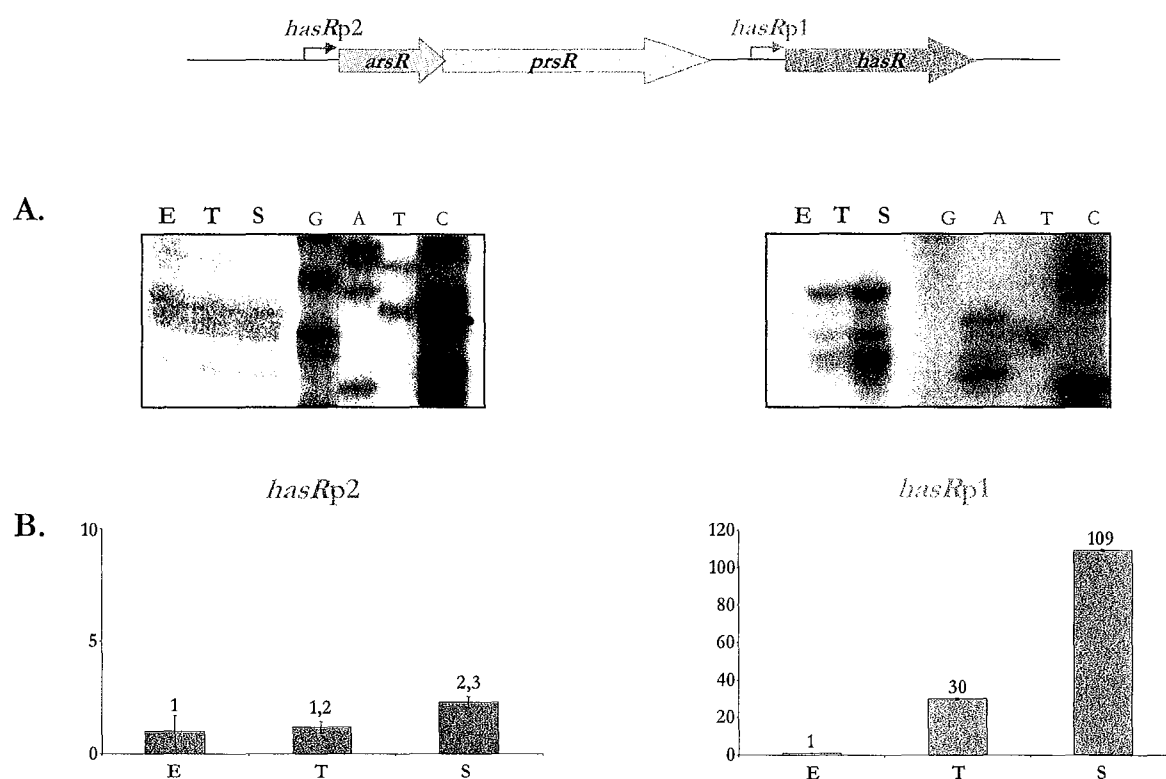


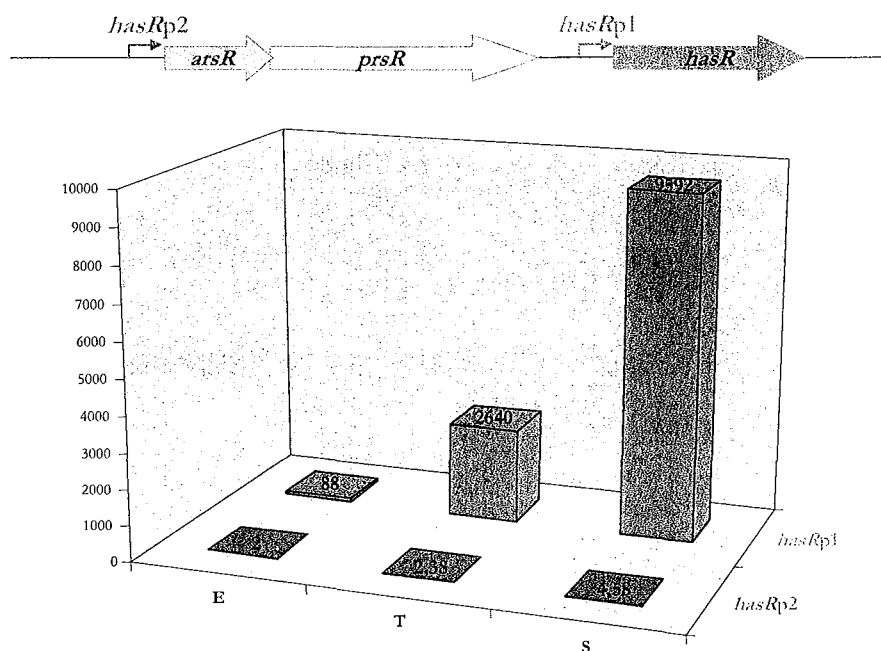
Figure R-59 : Régulation de l'expression du gène *hasR* au cours de la croissance.

A. Rappels des résultats des expériences de cartographie à la nucléase S1.

B. Quantification des transcrits initiés aux sites *hasRp2* et *hasRp1* dans les trois phases de croissance (E, exponentielle ; T, transition ; S, stationnaire) par RT-PCR quantitative. Le niveau d'expression de chaque transcrit est arbitrairement fixé à 1 en phase exponentielle de croissance. Il est à noter que les échelles des deux graphiques sont différentes.

B-III.2.2.2.2.2 – Abondance relative des transcrits initiés aux sites *hasRp1* et *hasRp2*

Enfin, ces données quantitatives révèlent l'abondance relative de chacun des deux transcrits du gène *hasR* et montrent, comme le suggéraient les résultats de cartographie à la nucléase S1, une très faible représentation du transcrit initié au site *hasRp2* comparativement à celui initié au site *hasRp1* (avec un niveau de transcription jusqu'à 2085 fois plus élevé en phase stationnaire de croissance, Tableaux R-8 et R-9 et Figure R-60).



	E	T	S
■ <i>hasRp2</i>	2 ± 1.4	2.38 ± 0.5	4.58 ± 0.5
■ <i>hasRp1</i>	88 ± 14.2	2640 ± 38	9592 ± 45.6

**Figure R-60 :** Abondance relative de chacun des deux transcrits du gène *hasR* au cours de la croissance.

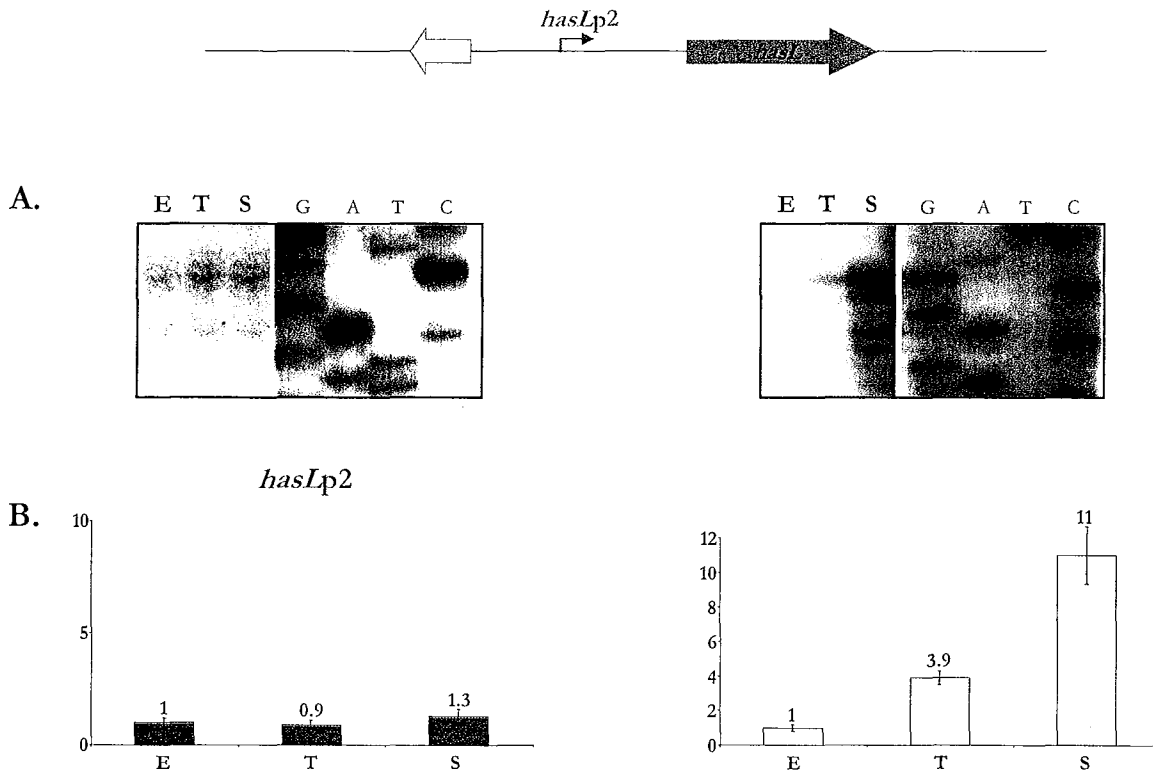
Dans cette représentation, le niveau d'expression globale du gène *hasR* (*hasRp1*+*hasRp2*) est arbitrairement fixé à 100 en phase exponentielle de croissance

B-III.2.2.2.3 – Quantification de la transcription du gène *hasL*B-III.2.2.2.3.1 – Evolution de la quantité de chaque transcrit du gène *hasL* au cours de la croissance

De façon similaire, l'analyse de la transcription du gène *hasL* par PCR quantitative a permis de confirmer que la transcription initiée au site *hasLp2* était constante (variation maximale d'un facteur 1,3), alors que celle initiée au site *hasLp1* était induite au cours de la croissance, avec une



induction d'un facteur 4 lors du passage de la phase exponentielle à la phase de transition et d'un facteur 11 lors du passage de la phase exponentielle à la phase stationnaire de croissance (Figure R-61 B.).



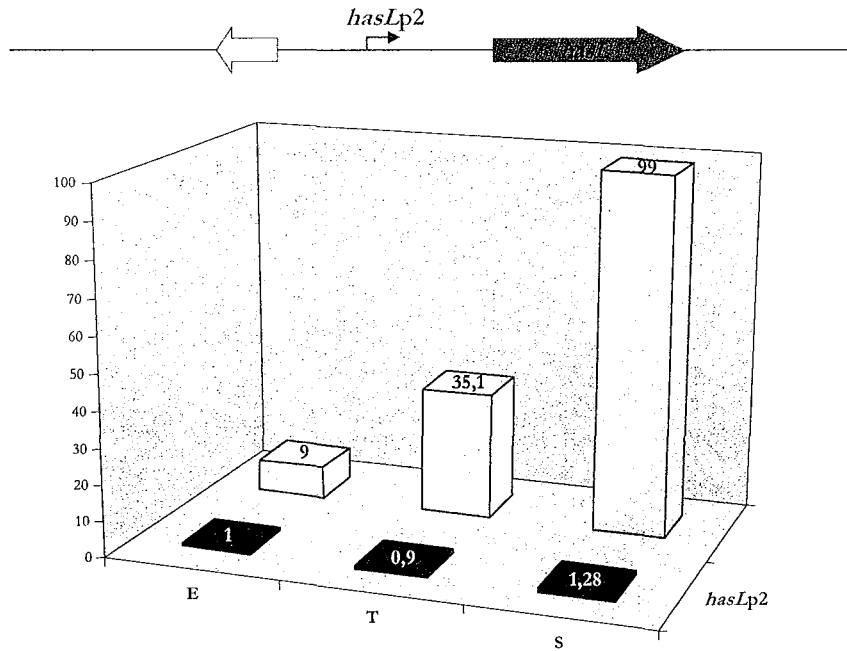
**Figure R-61 : Régulation de l'expression du gène *hasL* au cours de la croissance.**

A. Rappels des résultats des expériences de cartographie à la nucléase S1.

B. Quantification des transcrits initiés aux sites *hasLp2* et *hasLp1* dans les trois phases de croissance (E, exponentielle ; T, transition ; S, stationnaire) par RT-PCR quantitative. Le niveau d'expression de chaque transcrit est arbitrairement fixé à 1 en phase exponentielle de croissance. Il est à noter que les échelles des deux graphiques sont différentes.

B-III.2.2.2.3.2 – Abondance relative des transcrits initiés aux sites *hasLp1* et *hasLp2*

Comme pour le gène *hasR*, la transcription du gène *hasL* la plus abondante est celle initiée au site *hasLp1* (avec un niveau de transcription jusqu'à 76 fois plus élevé en phase stationnaire de croissance, Tableaux R-8 et R-9 et Figure R-62).



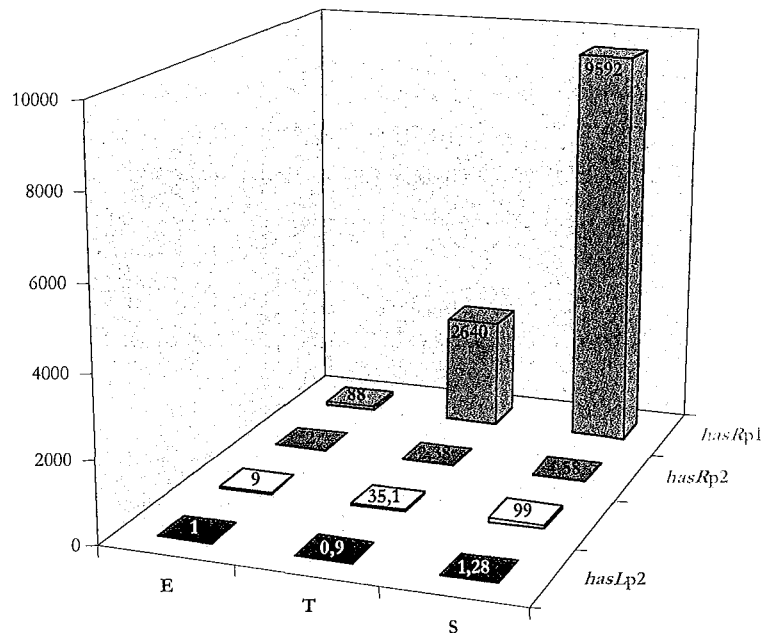
	E	T	S
■ <i>hasLp1</i>	1 ± 0.2	0.9 ± 0.2	1.28 ± 0.3
□ <i>hasLp2</i>	9 ± 1.6	35.1 ± 3.6	99 ± 14.9

Figure R-62 : Abondance relative de chacun des deux transcrits du gène *hasL* au cours de la croissance.

Dans cette représentation, le niveau d'expression globale du gène *hasL* (*hasLp1*+*hasLp2*) est arbitrairement fixé à 100 en phase exponentielle de croissance

**B-III.2.2.2.4 – Expression relative des gènes *hasR* et *hasL* au cours de la croissance**

Enfin, cette analyse a également permis de comparer l'expression relative de chaque transcrit et ainsi de montrer que le gène *hasR* était largement plus transcrit que le gène *hasL*, puisque son niveau d'expression est près de 100 fois plus élevé en phase stationnaire de croissance (Tableaux R-8 et R-9 et Figure R-63).



	E	T	S
■ <i>hasLp2</i>	1 ± 0.2	0.9 ± 0.2	1.28 ± 0.3
□	9 ± 1.6	35.1 ± 3.6	99 ± 14.9
▣ <i>hasRp2</i>	2 ± 1.4	2.38 ± 0.5	4.58 ± 0.5
▤ <i>hasRp1</i>	88 ± 14.2	2640 ± 38	9592 ± 45.6

**Figure R-63 : Abondance relative de chacun des quatre transcrits des gènes *has* au cours de la croissance.**

Dans cette représentation, le niveau d'expression globale des gènes *has* (*hasRp1* + *hasRp2* + *hasLp1* + *hasLp2*) est arbitrairement fixé à 100 en phase exponentielle de croissance

## B-III.3 $\propto$ ANALYSE DE L'EXPRESSION SPATIO-TEMPORELLE DES PROMOTEURS *HAS*

### III.3.1 - UTILISATION DU GÈNE RAPPORTEUR CODANT LA RÉSISTANCE A LA KANAMYCINE

Une première approche fonctionnelle a permis de rechercher et tester l'activité promotrice des régions 5' des gènes *has* par leur clonage en amont du gène *neo*. En effet, la force d'un promoteur putatif peut être mesurée par le niveau de résistance à la kanamycine conférée à une souche hôte contenant la construction.

#### B-III.3.1.1 - CLONAGE DES RÉGIONS PROMOTRICES DES GÈNES *HASR* ET *HASL*

Une ligation a été réalisée entre des digestions *SacI* de l'ADN du plasmide pIJ487 (voir Matériel et Méthodes) et des ADN des bactériophages M13 recombinants pI329 et pI338 contenant les régions promotrices des gènes *hasL* et *hasR* respectivement, sous forme d'inserts *SacI* (2,35 kb et 2,1 kb ; Figure R-64).

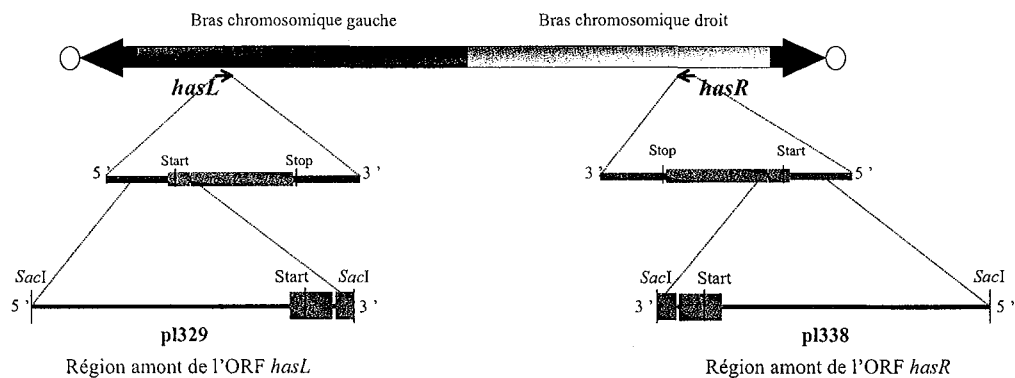


Figure R-64 : Représentation de la localisation chromosomique des ORF *has* et des inserts des phages M13 recombinants pI329 et pI338 contenant les régions amont des ORF *has* chez la souche sauvage *S. ambifaciens* DSM40697.

Les régions amont des ORF sont contenues dans des inserts *SacI* clonés dans le bactériophage M13 (Fischer *et al.*, 1998b). La flèche ( $\rightarrow$  ou  $\leftarrow$ ) indique l'orientation transcriptionnelle des ORF. Les séquences identiques à 100% entre les loci *hasR* et *hasL* sont représentées en bleu.

Ce mélange de ligation a ensuite été utilisé pour transformer des protoplastes de *S. lividans* ZX7 (dérivée de la souche sauvage *S. lividans* 1326), qui présente une déficience pour la recombinaison et pour les systèmes de restriction/modification permettant ainsi le clonage favorable de l'ADN de *Streptomyces* (voir Matériel et Méthodes).

Après sélection des transformants sur leur résistance au thiostrepton, une analyse de restriction par l'endonucléase *AseI* a permis de distinguer les souches renfermant un plasmide recombinant de celles contenant le vecteur de clonage non recombiné. En effet, le vecteur pIJ487 possède un site unique *AseI* alors que l'insert n'en contient pas. Ainsi, un vecteur non recombiné sera observé sous forme d'une bande de 6,2 kb, taille du vecteur pIJ487, alors qu'un recombinant sera observé sous forme d'une bande 8,3 kb ou 8,55 kb s'il contient l'insert de pl338 et pl329 respectivement.

Pour le clonage de la région 5' du locus *basR* (fragment *SacI* de pl338), 22 transformants ont été repiqués et leur ADN plasmidique extrait. Leur profil de restriction *AseI* présente une bande unique de 6,2 kb signifiant que les plasmides ne renferment pas d'insert.

Pour le clonage de la région 5' du locus *basL* (fragment *SacI* de pl329), 19 des 203 transformants ont été isolés et leur ADN plasmidique analysé. La digestion *AseI* montre que 2 des 19 plasmides présentent une bande aux environs de 8,5 kb suggérant qu'ils sont recombinés. En revanche, les 17 autres présentent une bande à 6,2 kb montrant qu'ils ne contiennent pas d'insert. L'analyse de restriction par *SacI* révèle deux bandes distinctes dans le cas des plasmides recombinés : une bande d'environ 6,2 kb, correspondant à pIJ487 et une bande d'environ 2,3 kb, correspondant à la taille de l'insert *SacI* de pl329. Ces deux plasmides recombinés sont appelés pNSL4 et pNSL9, et les souches de *S. lividans* ZX7 les renfermant, *S. lividans* NSL4 et NSL9 pour Nancy *Streptomyces lividans*. Un exemple de chaque type plasmidique est présenté sur la Figure R-65.

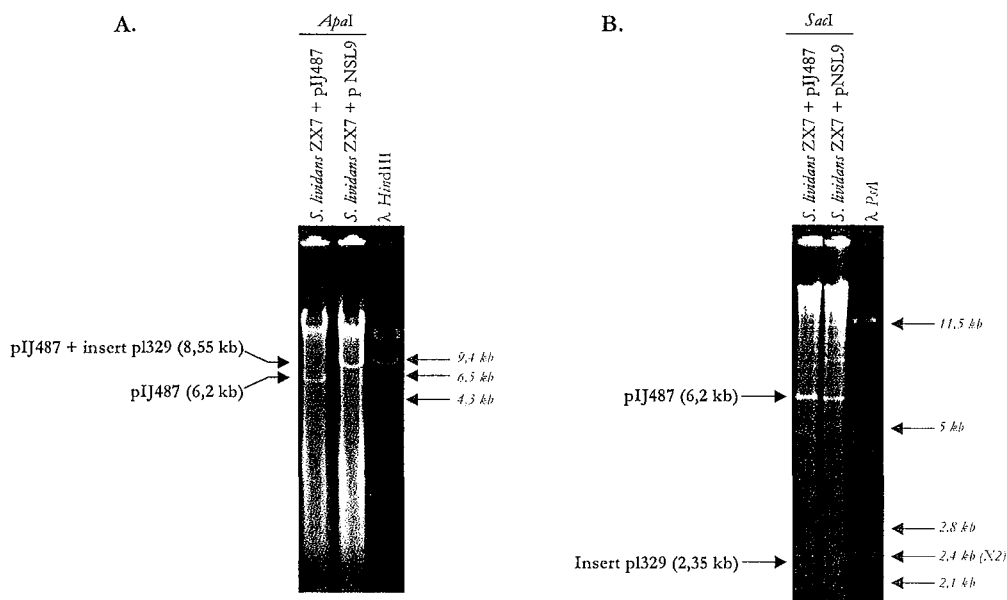


Figure R-65 : Comigrations électrophorétiques des profils de restriction *AseI* (A) et *SacI* (B) des ADN plasmidiques pIJ487 et pNSL9.

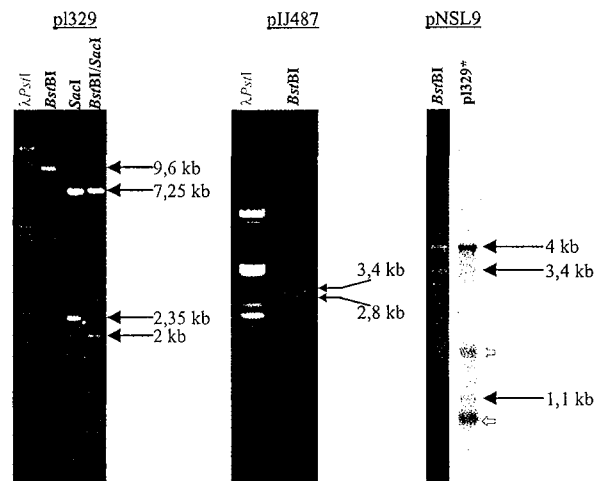
Les conditions de migration sont indiquées dans le matériel et méthodes. Seules les tailles indicatives des bandes du marqueur de taille ( $\lambda$ PstI) sont notées.

B-III.3.1.2 - CARTOGRAPHIE DE RESTRICTION DU PLASMIDE pNSL9

Avant de tester la présence d'une force promotrice en amont du gène *basL*, l'orientation de l'insert par rapport au gène *néo* dans les plasmides pNSL4 et pNSL9 a été déterminée. En effet, la région 5' du locus *basL* et le gène rapporteur *néo* doivent former une fusion transcriptionnelle pour conférer la résistance à la kanamycine. Deux endonucléases de restriction, *BstBI* et *ApaI* ont été utilisées à cette fin.

B-III.3.1.2.1 - Cartographie de restriction *BstBI* du plasmide pNSL9

L'insert *basL* (2,35 kb) possède un site unique *BstBI*. Ce site a été décelé lors du séquençage partiel de la région 5' du gène *basL* et par la digestion du bactériophage recombinant pI329 par *BstBI*. Cette digestion génère un fragment unique d'une taille de 9,6 kb, équivalente à la somme de l'insert (2,35 kb) et du vecteur M13 (7,25 kb) (Figure R-66).



**Figure R-66 : Cartographie *BstBI* du plasmide recombiné pNSL9.**

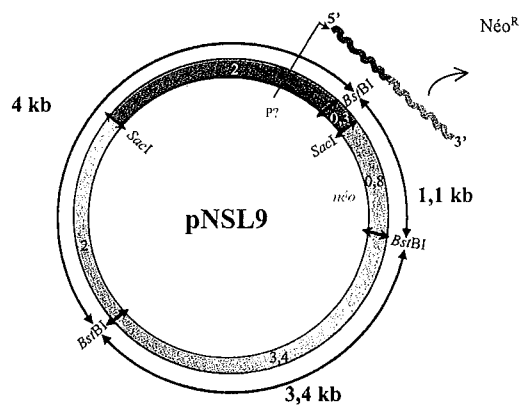
Digestions *BstBI* et/ou *SacI* du bactériophage pI329, des plasmides pIJ487 et pNSL9 et hybridation de l'ADN de pI329 marqué sur le profil de restriction *BstBI* de pNSL9. Les conditions d'électrophorèse et d'hybridation sont décrites dans le matériel et méthodes. Les signaux indiqués par les flèches ( $\leftrightarrow$ ) correspondent à des bandes non visibles sur le gel, donc ne révélant pas des fragments *BstBI*.

De plus, l'ADN du bactériophage M13 ne contient pas de site *BstBI* (non montré). La séquence de la région en 5' du gène *basL* révèle également la position du site *BstBI* : il est situé à 318 pb du site *SacI*, flanquant l'insert côté 3'. Cette localisation est corroborée par la double digestion *SacI*-*BstBI* du phage pI329 en comparaison de la digestion simple *SacI*. Cette dernière libère l'insert, 2,35 kb, et le vecteur, 7,25 kb (Figure R-66). L'endonucléase *BstBI* clive l'insert et génère un fragment d'une taille d'environ 2 kb (Figure R-66). Ainsi, cette disposition asymétrique du site *BstBI* dans l'insert peut être mise à profit pour déterminer l'orientation de l'insert.

Les plasmides pNSL4 et pNSL9 sont des dérivés de pIJ487. Le plasmide pIJ487 possède deux sites *BstBI* ; une digestion *BstBI* génère deux fragments de 2,8 kb et 3,4 kb (Figure R-66). Le site

de clonage *SacI* est localisé à l'intérieur du fragment de 2,8 kb, à 2 kb et 0,8 kb des sites *BstBI* (Figure R-67). Les profils de digestion *BstBI* des deux plasmides recombinants pNSL4 et pNSL9 sont identiques et présentent 3 fragments de taille 4 kb, 3,4 kb et 1,1 kb (Figure R-66). Les plasmides portent donc l'insert dans la même orientation et seul pNSL9 a été analysé plus avant. L'orientation de l'insert peut être déduite de la seule analyse du profil de restriction. En effet, le fragment de 3,4 kb provient du vecteur pIJ487, fragment non coupé par *SacI*. Compte tenu de la position du site *BstBI* dans l'insert, le fragment de 1,1 kb est constitué du fragment de 318 pb de l'insert et du fragment de 0,8 kb du vecteur. Le fragment de 4 kb résulte de la fusion des fragments de 2 kb de l'insert (précisément 2050 pb) et de 2 kb du vecteur.

Ces données sont confirmées par l'hybridation de l'ADN marqué de pI329 sur le profil de restriction *BstBI* de pNSL9 (Figure R-66). Le fragment de 3,4 kb n'est pas révélé car il ne possède que des séquences du vecteur pIJ487 (pIJ487 ne présente aucune homologie ni avec le bactériophage M13 ni avec l'insert, résultats non montrés). En revanche, les deux fragments de 4,1 kb et 1,2 kb sont révélés, donc possèdent des séquences homologues à l'insert. L'intensité des signaux est cohérente avec la longueur d'homologie avec l'insert ; 300 pb pour le fragment de 1,2 kb et 2 kb pour le fragment de 4,1 kb. L'orientation par rapport au gène *néo* est schématisée sur la Figure R-67.



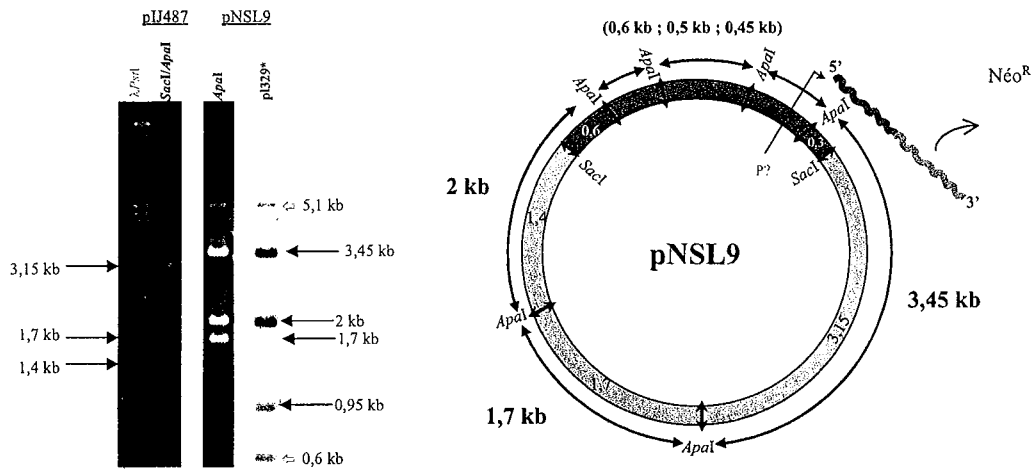
**Figure R-67 : Carte de restriction *BstBI* / *SacI* de pNSL9.**

L'insert est représenté en vert et le vecteur en orange. Les flèches bleues indiquent les tailles des fragments *BstBI* de pNSL9. La taille des fragments est indiquée en kilobases. P? : promoteur de transcription potentiel.  $\sim$  : transcrit putatif de la fusion transcriptionnelle.  $\text{Né}^{\text{R}}$  : représente l'aminoside phosphotransférase (NéoR) conférant la résistance à la kanamycine.

#### **B-III.3.1.2.2 - Cartographie de restriction *ApaI* du vecteur recombiné pNSL9**

Ces interprétations ont été confirmées par l'analyse par des profils de restriction obtenus après digestion par l'endonucléase *ApaI* et par l'hybridation de l'ADN marqué de pI329 sur le profil de restriction *ApaI* de pNSL9 (Figure R-68).

Ces résultats permettent de conclure quant à l'orientation de l'insert dans pIJ487, à savoir dans le sens de la fusion transcriptionnelle avec le gène *néo*. La carte *ApaI-SacI* du plasmide recombinant pNSL9 est donnée en Figure R-68.



**Figure R-68 : Cartographie *Apal* / *SacI* du plasmide recombiné pNSL9.**

A gauche : Digestions *Apal* des plasmides pIJ487 et pNSL9 et hybridation de l'ADN de p1329 marqué sur le profil de restriction *Apal* de pNSL9. Les flèches ( $\rightleftharpoons$ ) indiquent des bandes de digestion partielle. Les conditions d'électrophorèse et d'hybridation sont décrites dans le matériel et méthodes.

A droite : Carte de restriction *Apal* / *SacI* de pNSL9. La taille des fragments est indiquée en kilobases. Les fragments indiqués entre parenthèses sont non ordonnés. Le reste du symbolisme utilisé est le même qu'en Figure R-67.

### B-III.3.1.3 - DETECTION D'UNE FORCE PROMOTRICE EN AMONT DE HASL

#### B-III.3.1.3.1 - Expression de la résistance à la kanamycine chez *S. lividans* NSL9 en croissance sur milieu solide

L'expression de la résistance à la kanamycine a été testée sur milieu solide. En effet, ces conditions de croissance permettent la complète différenciation des *Streptomyces* (mycélium végétatif, mycélium aérien et spores) et permettent ainsi de mimer les conditions naturelles de croissance de *S. lividans*. Ainsi, deux souches sont testées : la souche *S. lividans* NSL9 et une souche témoin *S. lividans* ZX7 transformée par le plasmide pIJ487. Des spores des 2 souches sont étalées sur milieu HT contenant des doses croissantes de kanamycine variant de 0 à 500  $\mu\text{g/ml}$ , à raison de  $10^8$  UFC par boîte de façon à observer la formation de tapis confluents.

Après trois jours de culture à 30°C, aucune croissance n'est observée pour la souche témoin quelle que soit la dose de kanamycine alors qu'un tapis confluent est présent sur milieu HT sans antibiotique. En revanche, un tapis confluent est observé pour la souche *S. lividans* NSL9 jusqu'à la concentration de 100  $\mu\text{g/ml}$  incluse.



Ainsi, la région 5' du locus *hasL*, clonée en amont du gène *neo* dans pNSL9, permet aux souches renfermant ce plasmide d'exprimer une résistance à la kanamycine et contient donc un ou plusieurs promoteurs conférant une résistance à des doses allant de 25 µg/ml à 100 µg/ml. De plus, la force promotrice permet l'expression du gène *neo* dès la germination, soit en début de phase exponentielle.

### III.3.1.3.2 - Expression de la résistance à la kanamycine chez *S. lividans* NSL9 au cours de la croissance en milieu liquide

Après établissement des courbes de croissance de la souche *S. lividans* NSL9 et de la souche témoin (*S. lividans* ZX7 transformée par pIJ487, voir Matériel et Méthodes) et pour les temps 19 h, 23 h et 41 h correspondant respectivement à la phase exponentielle, phase de transition et phase stationnaire, une série de suspensions dilutions des cultures est étalée sur milieu HT contenant des doses croissantes de kanamycine variant de 0 à 800 µg/ml. Après 3 jours de croissance à 30°C, les colonies sont comptées (Tableau R-10).

	Phase exponentielle		Phase de transition		Phase stationnaire	
	<i>S. lividans</i>	<i>S. lividans</i>	<i>S. lividans</i>	<i>S. lividans</i>	<i>S. lividans</i>	<i>S. lividans</i>
	ZX7 + pIJ487	NSL9	ZX7 + pIJ487	NSL9	ZX7 + pIJ487	NSL9
Dose de	25	0% 97,5 +/- 4,3%	0% 87,4 +/- 17,8%	0% 57,1 +/- 11,3%	0% 59,1 +/- 7,4%	0% 95,4 +/- 4,3%
kanamycine	50	0% 70,1 +/- 30,1%	0% 19,5 +/- 2,0%	0% 0,34 +/- 0,35%	0% 0%	0% 36,5 +/- 10,2%
(µg / ml)	100	0% 34,9 +/- 11,8%	0% 0%	0% 0%	0% 0%	0% 0%
	200	0% 0,92 +/- 2%	0% 0%	0% 0%	0% 0%	0% 0%
	600	0% 0%	0% 0%	0% 0%	0% 0%	0% 0%
	800	0% 0%	0% 0%	0% 0%	0% 0%	0% 0%

Tableau R-10 : Expression de la résistance à la kanamycine chez *S. lividans* NSL9 en croissance en milieu liquide.

Les pourcentages expriment le rapport entre le nombre d'unités formant colonies (UFC) obtenues sur milieu HT additionné d'une dose de kanamycine par rapport au nombre d'UFC obtenues par étalement de la même suspension/dilution sur milieu HT sans kanamycine.

La souche témoin *S. lividans* ZX7 transformée par le plasmide pIJ487 n'exprime la résistance à la kanamycine dans aucune des 3 phases de croissance. En revanche, la souche *S. lividans* NSL9 exprime la résistance jusqu'à une dose de 100 µg/ml incluse et ce, dans les 3 phases de croissance. Ainsi, la force promotrice semble constante au cours des phases de croissance

étudiées, dans la limite de la sensibilité de la méthode utilisée. En outre, la résistance conférée en phase exponentielle de croissance sur milieu solide et liquide est similaire, soulignant une activité promotrice comparable.

Par ailleurs, le pourcentage de la population se développant sur les milieux décroît de 100% à environ 25% proportionnellement à l'augmentation de la dose d'antibiotique. La taille moyenne des colonies observées diminue également en fonction de la dose. Ces résultats peuvent être interprétés en terme de dose subinhibitrice d'antibiotique ou de variation du nombre de copies du plasmide pNSL9 au sein de la population.

### B-III.3.2 - UTILISATION DU GÈNE RAPPORTEUR CODANT L'EGFP

Suite à l'identification des différents sites d'initiation de la transcription des gènes *has*, une seconde approche visant à étudier la fonctionnalité de ces promoteurs a été développée. Cette approche consiste à étudier l'expression d'un gène rapporteur codant une protéine fluorescente, l'EGFP (Enhanced Green Fluorescent Protein, (Sun *et al.*, 1999)), après avoir cloné en amont de celui-ci les régions promotrices des gènes *has*, limitées cette fois-ci à une région d'environ 200 pb en amont de chaque codon d'initiation de la traduction.

#### B-III.3.2.1 - CLONAGE DES REGIONS PROMOTRICES DES GENES *HASL* ET *HASR* DANS LE VECTEUR PIJ8630

Une réaction de polymérisation en chaîne (PCR) sur l'ADN simple brin (rendement supérieur à l'ADN double brin) issu du vecteur recombinant du phage M13 a permis d'amplifier, à partir des couples d'amorces phasL1/phasL2 et phasR1/phasR2, deux séquences d'ADN. Ces deux amplifiats ont une taille d'environ 225 et 235 pb et peuvent correspondre aux deux régions attendues de 227 et 233 pb situées en amont et chevauchant de quelques nucléotides la phase codante des gènes *hasL* et *hasR* respectivement (Figure R-69).

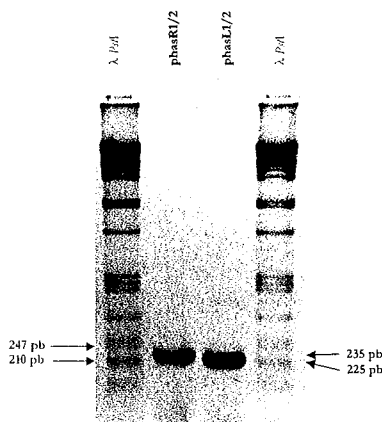


Figure R-69 : Migration électrophorétique des produits de PCR phasL1/2 et phasR1/2.

Les conditions de migration sont indiquées dans le matériels et méthodes. Seules les tailles des bandes du marqueur de taille  $\lambda$ PstI (en gris) encadrant les amplifiats (taille des bandes pointées par les flèches noires) sont indiquées. Des quantités équivalentes d'ADN amplifié ont été déposés.

Les étapes qui suivent sont réalisées en parallèle sur l'ADN amplifié de la région promotrice de *hasL* (PhasL) et *hasR* (PhasR).

Une ligation est ensuite réalisée entre ces produits de PCR, digérés par l'endonucléase *KpnI*, et le plasmide intégratif pIJ8630 (voir Matériel et Méthodes), lui même digéré par *KpnI* et déphosphorylé.

Ces deux mélanges de ligation ont été utilisés pour transformer des cellules compétentes d'*E. coli* SURE™ par électroporation (voir Matériel et Méthodes).

Afin d'identifier, parmi les clones transformants, ceux portant un vecteur recombinant, des digestions enzymatiques par l'enzyme *KpnI* de l'ADN plasmidique de clones pris au hasard ont été réalisées. Les plasmides ayant intégré la région promotrice du gène *hasR* ou *hasL* seront appelés pIJR ou pIJL respectivement et les souches d'*E. coli* ayant intégré ces vecteurs NECpIJR et NECpIJL respectivement (NEC pour Nancy *Escherichia coli*).

Pour la transformation avec le mélange de ligation renfermant PhasL, 18 clones (sur 136) ont été testés. Parmi ces 18 clones, 6 présentent un profil de restriction *KpnI* cohérent avec la présence d'un plasmide renfermant un insert PhasL (Figure R-60). En effet, les deux bandes caractéristiques (environ 215 pb et 8 kb) sont obtenues. Ces 6 clones possédant sont appelés NECpIJL5, NECpIJL6, NECpIJL7, NECpIJL11, NECpIJL17 et NECpIJL18.

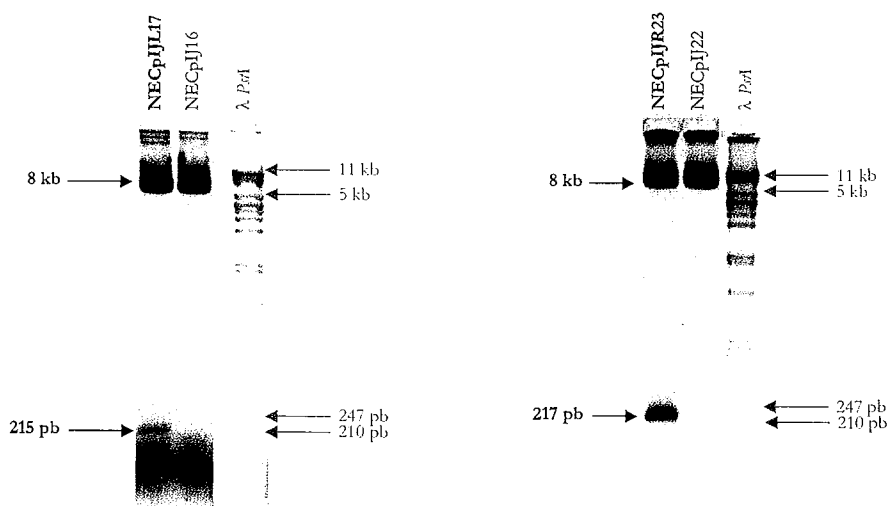


Figure R-70 : Migration électrophorétique de l'ADN plasmidique digéré par *KpnI* de deux clones transformés par le mélange de ligation pIJ8630 + PhasL (clones 16 et 17) et de deux clones transformés par le mélange de ligation pIJ8630 + PhasR (clones 22 et 23).

Les conditions de migration sont indiquées dans le matériel et méthodes. Seules les tailles des bandes du marqueur de taille  $\lambda PstI$  (en gris) encadrant les bandes d'intérêt (flèches noires) sont indiquées. Les noms de plasmides en gras correspondent aux clones recombinants.

Pour la transformation avec le mélange de ligation renfermant PhasR, 28 clones (sur 63) ont été testés et 7 sont apparus comme ayant intégré un plasmide recombinant (Figure R-70). En effet, la migration électrophorétique des digestions *KpnI* de l'ADN plasmidique des transformants montre 7 profils de restriction présentant une bande aux alentours de 8 kb (plasmide vide) et une bande

d'environ 220 pb (PhasR). Ils sont appelés NECpIJR6, NECpIJR12, NECpIJR18, NECpIJR21, NECpIJR23, NECpIJR26 et NECpIJR28.

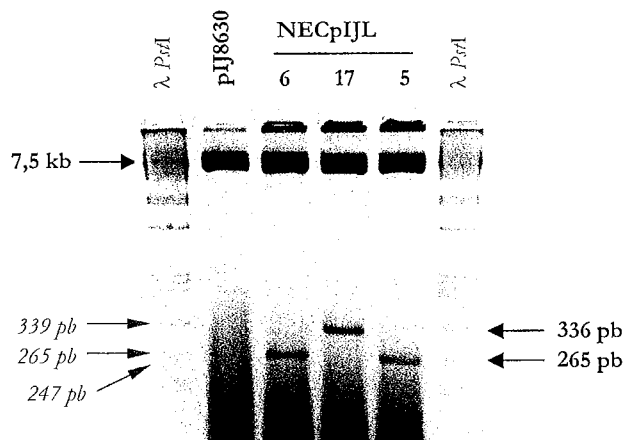
### B-III.3.2.2 - CARTOGRAPHIE DE RESTRICTION DES PLASMIDES PIJR ET PIJL

Les fragments de PCR ayant pu s'intégrer au site *KpnI* de la région MCS du plasmide pIJ8630 dans les deux sens par rapport au gène rapporteur et en tandem, des expériences utilisant des enzymes de restriction ont été menées afin de déterminer le nombre et l'orientation des inserts.

#### B-III.3.2.2.1 – Les vecteurs recombinants pIJL

Le nombre d'insert(s) au site de clonage *KpnI* a été déterminé par une double digestion enzymatique *BglII/XbaI*. En effet, le vecteur pIJ8630 possède un site unique pour chacune de ces deux enzymes, situés de part et d'autre du site de clonage *KpnI*.

Cette analyse de restriction concerne les clones NECpIJL5, NECpIJL6 et NECpIJL17 (Figure R-71). Les deux premiers clones présentent une bande d'environ 265 pb, fait cohérent pour une insertion unique. Le clone NECpIJL17 présente une bande de taille supérieure (environ 340 pb) mais ne correspond pas à une insertion en tandem. Les 80 pb supplémentaires pourraient provenir d'un fragment de PCR aspécifique non détecté du fait même de sa faible taille.

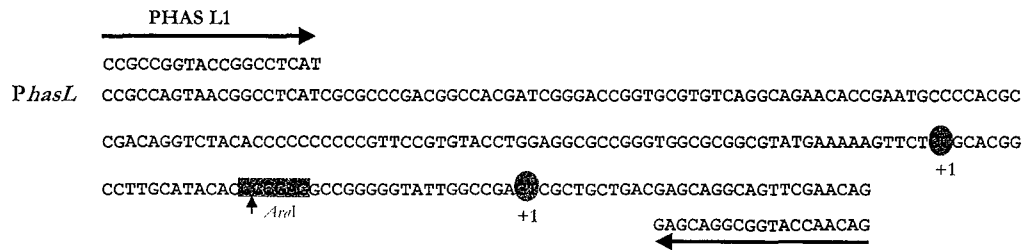


**Figure R-71 : Migration électrophorétique de l'ADN plasmidique digéré par *BglII / XbaI* des recombinants NECpIJL6, 17 et 5.**

Les conditions de migration sont indiquées dans le matériels et méthodes. Seules les tailles des bandes du marqueur de taille  $\lambda PstI$  (en gris) encadrant les bandes d'intérêt (flèches noires) sont indiquées.

Afin de déterminer l'orientation de l'insert, des double digestions enzymatiques *AvaI / XbaI* de l'ADN plasmidique des vecteurs recombinants ont été réalisées.

En effet, l'insert PhasL (227 pb) possède un site unique de coupure par l'endonucléase de restriction *AvaI* en position asymétrique (insert PhasL fragmenté en 54 pb et 173 pb ; Figure R-72).



**Figure R-72 :** Séquences des régions promotrices du gène *hasL*. Les séquences sont écrites dans le sens 5' vers 3'. Les sites d'initiation de transcription putatifs sont symbolisés par les ovales oranges et le signe +1. Le couple d'amorces (séquences précisées) utilisé pour amplifier la région promotrice est indiqué par des flèches rouges. Le site de restriction *KpnI* inséré dans les oligonucléotides est indiqué en rouge. Le site de restriction *AvaI*, utilisé pour déterminer l'orientation de l'insert, est surligné en vert. La flèche noire verticale indique le site de clivage par l'endonucléase.

Le plasmide pIJ8630 présente quant à lui un seul site de restriction *BglII* situé à 25 pb en amont (sens arbitraire de lecture donné par l'orientation transcriptionnelle du gène rapporteur) du site de clonage *KpnI* dans lequel l'insert a été cloné. De plus, une digestion du plasmide natif par *AvaI* a permis de détecter la présence d'au moins un site de coupure pour cette enzyme sur pIJ8630 (résultats non montrés).

Ainsi, dans le cas où l'insert est orienté dans le sens de la fusion transcriptionnelle en copie unique la double digestion *AvaI* / *BglII* de l'ADN des plasmides recombinants pIJL doit révéler deux bandes signatures (Figure R-73) :

- une bande de 198 pb correspondant à la somme des fragments de 173 pb (fragment *KpnI*-*AvaI* de PhasL) et 25 pb (fragment *BglII*-*KpnI* de pIJ8630),
- une bande de X pb correspondant au fragment *AvaI*-*AvaI* généré par les coupures dans le site *AvaI* de l'insert PhasL et dans le site *AvaI* du plasmide pIJ8630 dont la position est inconnue. La taille de cette bande est fonction de la position du site *AvaI* dans le plasmide pIJ8630.

Dans le cas contraire, la digestion par les endonucléases *AvaI* / *BglII* doit libérer (Figure R-73) :

- un fragment de 79 pb correspondant à la somme des fragments de 54 pb (fragment *KpnI*-*AvaI* de PhasL) et 25 pb (fragment *BglII*-*KpnI* de pIJ8630),
- un fragment de Y pb correspondant au fragment *AvaI*-*AvaI* généré par les coupures dans le site *AvaI* de l'insert PhasL et dans le site *AvaI* du plasmide pIJ8630 dont la position est inconnue. L'orientation inverse de l'insert par rapport au premier cas entraîne l'apparition de ce fragment de Y pb plus grand de 119 pb (soit 173 pb - 54 pb) par rapport au fragment de X pb.

L'analyse des profils de restriction montre que NECpIJL6 et NECpIJL11 ont intégré un plasmide recombinant possédant un insert orienté dans le sens de la fusion transcriptionnelle. En effet, la bande X  $\approx$  445 pb et la bande d'environ 198 pb (200 pb mesurées) sont présentes (Figure R-73). L'ADN plasmidique des transformants NECpIJL5, 7 et 18 présentent des profils de restriction cohérents avec un événement d'insertion de PhasL dans le sens inverse puisque seule la bande Y  $\approx$  570 pb est présente. La taille de cette bande est bien supérieure à celle de la bande X  $\approx$  445 pb et la différence de taille (125 pb) est proche des 119 pb attendues. La bande de 79 pb attendue n'est pas visualisable car elle est masquée par de l'ARN contaminant (Figure R-73).

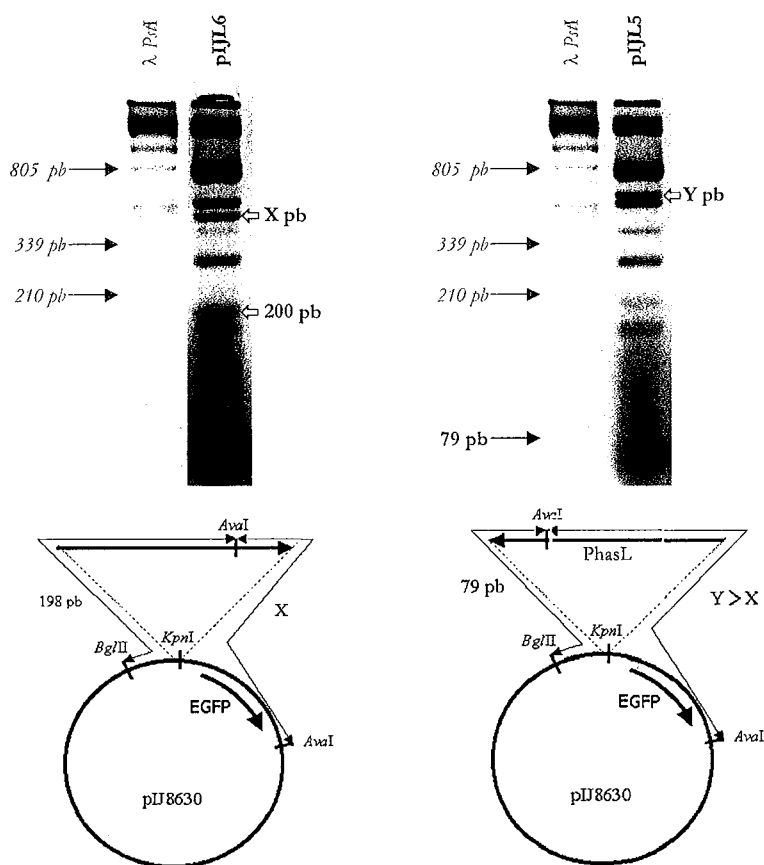


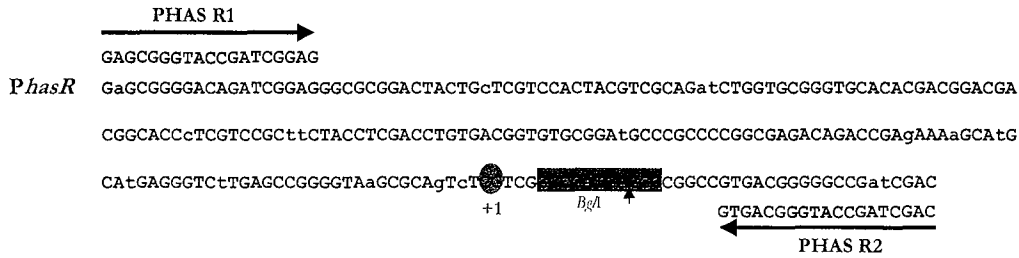
Figure R-73 : Détermination de l'orientation de l'insert PhasL dans les vecteurs recombinants pIJL6 et pIJL5.

En haut : Migration électrophorétique des digestions des plasmides pIJL6 et pIJL5 par *AvaI* et *BglII*. Les conditions de migration sont indiquées dans le matériels et méthodes. Seules les tailles des bandes du marqueur de taille  $\lambda PstI$  (en gris) encadrant les bandes sont indiquées. Les flèches rouges pointent les bandes caractéristiques.

En bas : Cartes de restriction de vecteurs recombinants où l'insert PhasL est inséré dans le sens de la fusion transcriptionnelle (à gauche, pIJL6) ou dans le sens contraire de la fusion transcriptionnelle (à droite, pIJL5).

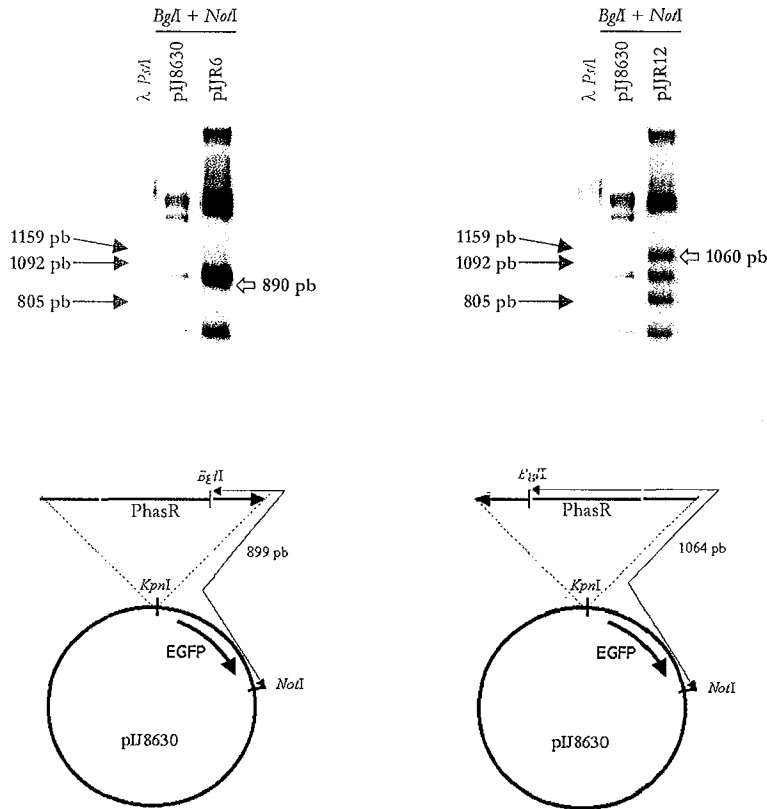
#### B-III.3.2.2.2 – Les vecteurs recombinants pIJR

De façon similaire, le nombre d'insert(s) PhasR ainsi que leur orientation ont été déterminés par des analyses de restriction (double digestions *BglII/NotI*, Figure R-74).



**Figure R-74 :** Séquences des régions promotrices du gène *hasR*. Le symbolisme est identique à celui de la Figure R-72. Le site de restriction *Bgl*I, utilisé pour déterminer l'orientation de l'insert, est surligné en bleu. La flèche noire verticale indique le site de clivage par l'endonucléase.

Par un raisonnement analogue, le transformant NECpIJR6 et 30 est interprété comme ayant intégré un plasmide recombinant possédant un insert unique orienté dans le sens de la fusion transcriptionnelle et les transformants NECpIJR12, 18 et 26 sont interprétés comme ayant intégré un plasmide recombinant dans lequel PhasR serait en copie unique et en orientation contraire à la fusion transcriptionnelle (Figure R-75).



**Figure R-75 :** Détermination de l'orientation de l'insert PhasR dans les vecteurs recombinants pIJR6 et pIJR12.

Le symbolisme est identique à celui de la Figure R-73.

En haut : Migration électrophorétique des digestions des plasmides pIJR6 et pIJR12 par *Bgl*I et *Not*I.

En bas : Cartes de restriction de vecteurs recombinants où l'insert PhasR est inséré dans le sens de la fusion transcriptionnelle (à gauche, pIJR6) ou dans le sens contraire de la fusion transcriptionnelle (à droite, pIJR12).

Par conséquent, NECpIJL6, NECpIJL11, NECpIJR6 et NECpIJR30 sont choisis pour représenter des événements de type insertion unique et dans le sens de la fusion transcriptionnelle. De même, deux clones transformants seront analysés pour chaque région promotrice des gènes *hasL* et *hasR* insérée en copie unique et en orientation inverse par rapport à la fusion transcriptionnelle, à savoir, NECpIJL5 et 7 et NECpIJR12 et 18.

### B-III.3.2.3 - ANALYSE DE LA FLUORESCENCE CHEZ *S. AMBOFACIENS*

Huit plasmides recombinants (NECpIJL6, NECpIJL11, NECpIJR6 et NECpIJR30 pour la fusion transcriptionnelle et NECpIJL5, NECpIJL7, NECpIJR12 et NECpIJR18 pour l'orientation inverse par rapport à la fusion transcriptionnelle) ont ensuite été transférés dans la souche *E. coli* S17, portant les fonctions nécessaires à la conjugaison intergénérique.

Après vérification des cartes de restriction des vecteurs ainsi obtenus dans la souche *E. coli* S17, huit souches donneuses ont été utilisées dans des expériences de conjugaison entre *E. coli* et *S. ambofaciens*. Les souches de *S. ambofaciens* ainsi transformées, sélectionnées sur leur résistance à l'apramycine conférée par la présence du plasmide, sont nommés NSApIJR ou NSApIJL suivi d'un numéro en fonction du vecteur intégré (NSA pour Nancy *S. ambofaciens*).

La visualisation de la fluorescence sur les préparations des transconjugants de *S. ambofaciens* reste cependant à mettre au point. En effet, deux expériences consécutives ont été réalisées, et alors que la fluorescence apparaissait comme significative lors de la première expérience, une auto-fluorescence, masquant tout autre signal, a été détectée lors des deuxièmes prises de vue.

Lors de la première expérience, trois préparations ont pu être analysées. Tout d'abord, sur la préparation correspondant à un exconjugant de *S. ambofaciens* ayant intégré un vecteur pIJ8630 vide, aucune fluorescence n'a pu être détectée. Seul un léger bruit de fond, ne correspondant pas à la localisation du matériel, a été observé (résultats non montrés).

Ensuite, la préparation de l'exconjugant de *S. ambofaciens* NSApIJL7, ayant intégré le vecteur pIJL7 renfermant la région promotrice du gène *hasL* insérée dans le sens contraire par rapport à la fusion transcriptionnelle, a été analysée. Sur cette préparation, un bruit de fond similaire à celui observé dans le cas du vecteur vide a été observé, démontrant ainsi l'absence d'expression de l'EGFP et donc l'absence d'initiation de la transcription du gène codant l'EGFP (Figure R-76).

Enfin, une analyse a également pu être effectuée sur la préparation de l'exconjugant de *S. ambofaciens* NSApIJL11, ayant intégré le vecteur pIJL11 renfermant la région promotrice du gène *hasL* insérée dans le sens de la fusion transcriptionnelle. Dans ce cas, l'émission d'une fluorescence (différente du bruit de fond) a pu être observée au niveau d'une chaînette de spores traduisant ainsi l'expression de l'EGFP dans ce compartiment (Figure R-76). Par ailleurs, une fluorescence de plus faible intensité a été observée dans un fragment mycélien.

Dans ces conditions, la région promotrice du gène *hasL* apparaît donc comme étant fonctionnelle. Le gène *hasL* pourrait donc être exprimé chez la souche sauvage de *Streptomyces ambofaciens* au niveau du mycélium végétatif ainsi qu'au niveau des spores.



La mise au point de la visualisation de la fluorescence dans les différentes constructions est actuellement en cours.

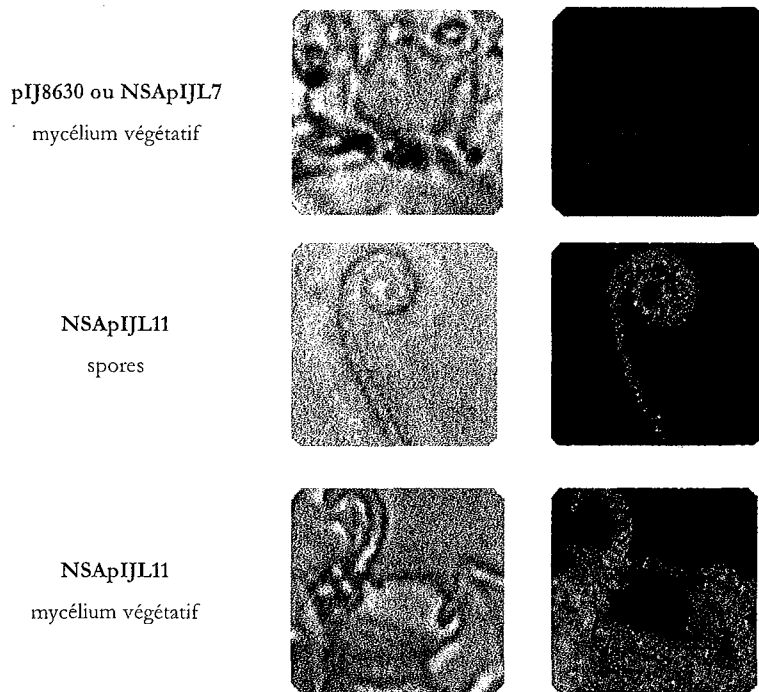


Figure R-76 : Analyse de la fluorescence, révélant l'activité de la région promotrice du gène *hasL*, chez *S. ambofaciens*.

- CHAPITRE B-IV -

CONSTRUCTION DE MUTANTS  
INTERROMPUS POUR LE(S) GENE(S)  
*HAS*

L'analyse fonctionnelle systématique impose une étape d'inactivation des ORF considérées. Chez les bactéries, les méthodes de remplacement de gènes nécessitent le plus fréquemment la création de l'interruption sur un plasmide adapté avant remplacement du locus chromosomique.

Par opposition, chez la levure *S. cerevisiae*, la recombinaison homologe est particulièrement efficace et les ORF sont interrompues directement par remplacement au locus chromosomique après transformation avec un ADN linéaire portant l'allèle muté (généralement produit par polymérisation en chaîne ou PCR pour "Polymerase Chain Reaction"). Cet ADN porte un marqueur sélectionnable (gène de résistance à un antibiotique) flanqué par des régions de taille aussi réduite que 35 nucléotides et homologues à la séquence chromosomique. Cette approche permet l'analyse fonctionnelle systématique de l'ensemble des ORF déduites de la séquence génomique de la levure (Baudin *et al.*, 1993).

Contrairement à la levure et à quelques espèces bactériennes naturellement compétentes, la plupart des bactéries ne sont pas transformables par des molécules d'ADN linéaires en raison de la présence du complexe exonucléasique RecBCD (ou exoV). En effet, cette activité réduit le temps de demi-vie de molécules d'ADN linéaire de façon drastique (Lorenz and Wackernagel, 1994). Toutefois, chez *E. coli*, l'utilisation de souches déficientes dans le système RecBCD a permis de contourner cette limitation.

Enfin, Murphy (Murphy, 1998) a montré que l'utilisation du système de recombinaison RED du phage  $\lambda$  permet d'augmenter chez la souche sauvage d'*E. coli*, le taux de recombinaison de l'ADN linéaire d'au moins 70 fois. Ce système a été utilisé pour promouvoir le remplacement de gènes chez *E. coli*.

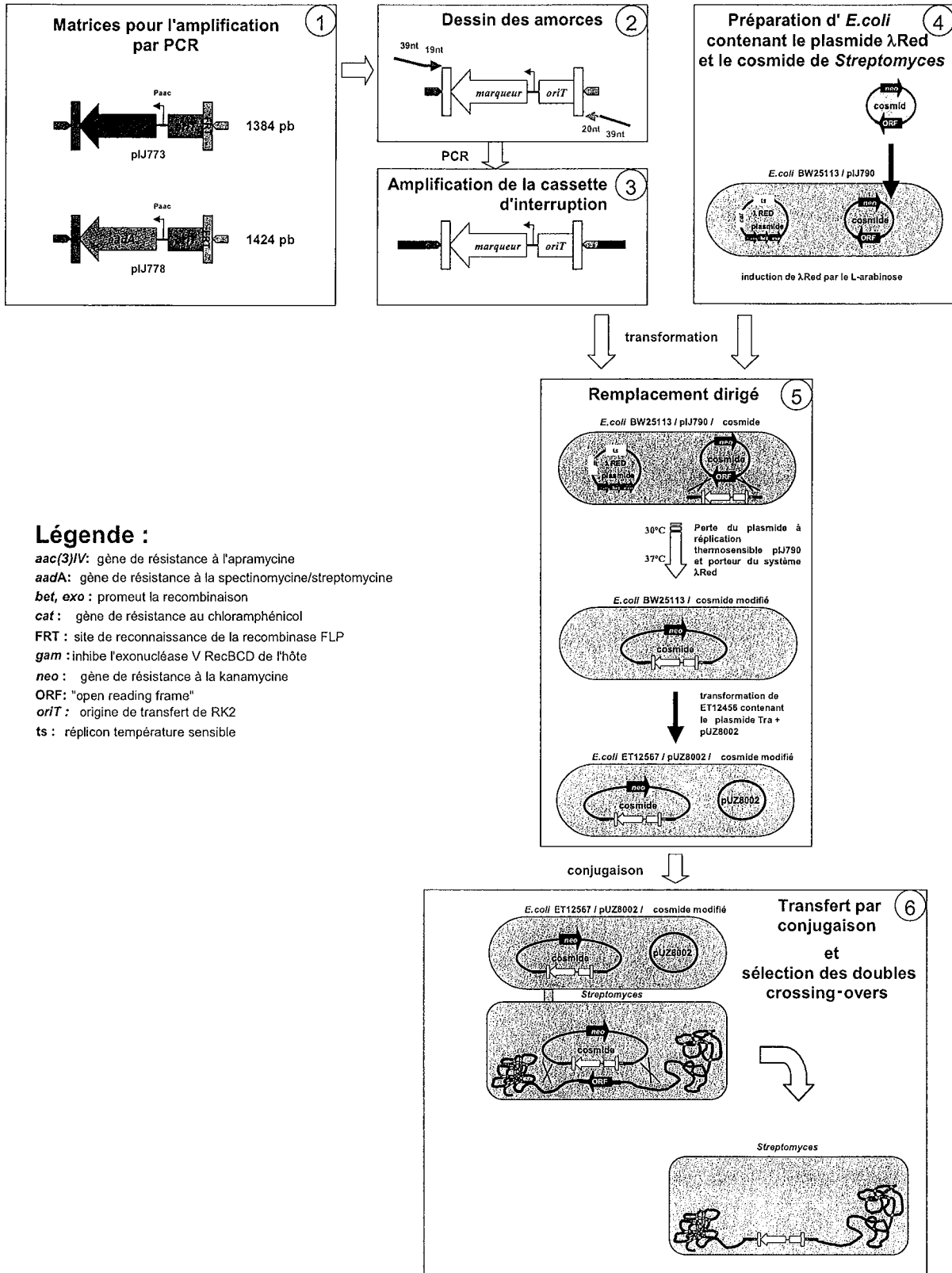
Le système  $\lambda$ RED est composé de 3 gènes, *gam-exo-bet*. La protéine Gam inhibe l'activité exonucléasique RecBCD, et la protéine Exo dégrade les molécules d'ADN double brin dans le sens 5' vers 3'. La protéine Bet se lie aux extrémités des molécules d'ADN simple brin, favorise l'hybridation d'extrémités complémentaires et stimule la recombinaison RecA dépendante.

En se basant sur ce système, Datsenko et Wanner (Datsenko and Wanner, 2000) ont développé une méthode simple et efficace pour interrompre des gènes chromosomiques chez *E. coli* dans laquelle des amorces fournissent le substrat de la recombinaison au locus considéré. Ces amorces servent à générer par PCR un ADN linéaire contenant une cassette de résistance et permettant le remplacement au locus chromosomique.

Cette méthode a été très récemment adaptée chez *S. coelicolor* A3(2) par Gust *et al.* (2003)

## B-IV.1 $\propto$ STRATEGIE DE REMPLACEMENT D'UNE ORF AU LOCUS CHROMOSOMIQUE CHEZ *STREPTOMYCES*

La stratégie se divise en trois étapes majeures schématisées sur la Figure R-77. La construction est réalisée à partir d'un cosmide recombinant chez l'hôte *E. coli*, puis transférée chez la souche de *Streptomyces* étudiée.



**Légende :**

- aac(3)/IV*: gène de résistance à l'apramycine
- aadA*: gène de résistance à la spectinomycine/streptomycine
- bet, exo*: promeut la recombinaison
- cat*: gène de résistance au chloramphénicol
- FRT: site de reconnaissance de la recombinase FLP
- gam*: inhibe l'exonucléase V RecBCD de l'hôte
- neo*: gène de résistance à la kanamycine
- ORF: "open reading frame"
- oriT*: origine de transfert de RK2
- ts: réplicon température sensible

Figure R-77 : Stratégie de remplacement de gène par "PCR targeting" (d'après Gust *et al.* (2003))

### B-IV.1.1 – CHOIX DES AMORCES ET FABRICATION DU PRODUIT DE PCR (ETAPE 1)

L'allèle muté est construit sur un cosmide recombinant (dérivé de Supercos1, (Evans *et al.*, 1989) recouvrant le locus considéré. Deux amorces de grande taille (58 et 59 nucléotides) sont nécessaires (Figure R-77, cadre 2). Chacune des amorces se caractérise par deux régions distinctes : une séquence de 39 nucléotides, située en 5' de l'amorce, est complémentaire des régions adjacentes du gène à remplacer, et en 3' une séquence de 19 ou 20 nucléotides est complémentaire des extrémités 5' ou 3' de la cassette (Figure R-78). L'amorce 5' ou « forward » inclut le codon d'initiation de traduction de l'ORF considérée tandis que l'amorce 3' ou « reverse » inclut le codon d'arrêt de la traduction. Ainsi, la cassette peut être orientée en combinant, au sein des amorces, les séquences homologues de la cassette et du locus considéré. Enfin, les séquences aux extrémités 5' et 3' des différentes cassettes disponibles sont identiques permettant ainsi d'amplifier chacune d'elles avec le même couple d'amorces.

Ces cassettes, outre le marqueur génétique, comprennent l'origine de transfert *oriT* du plasmide RK2 (Pansegrau *et al.*, 1994) et deux séquences FRT (pour "FLP Recognition Target"), sites de reconnaissance pour la recombinase site-spécifique FLP (Figure R-77, cadre 1). La présence des sites FRT permet, après remplacement de l'ORF étudiée, l'excision par la protéine FLP de la partie centrale de la cassette, si besoin est. Selon l'orientation de la cassette après intégration, l'excision aboutit soit à une délétion en phase de l'ORF, soit à l'introduction de codons d'arrêt de la traduction dans les trois phases de lecture.

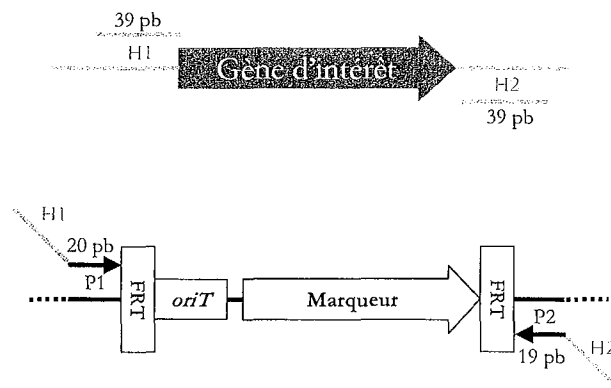


Figure R-78 : Dessin des amorces d'amplification des cassettes.

Les parties H1 et H2 (39 pb) des amorces sont complémentaires aux séquences flanquant le gène à interrompre. Les parties P1 (20 pb) et P2 (19 pb) sont complémentaires aux séquences flanquant la cassette.

### B-IV.1.2 – CONSTRUCTION DE L'ALLELE MODIFIE SUR LE COSMIDE RECOMBINANT (ETAPE 2)

Le cosmide recombinant recouvrant le locus d'intérêt est introduit par électroporation dans la souche *E. coli* BW25113/pIJ790 (Figure R-77, cadre 4). Ces cosmides sont issus d'une banque

génomique de *S. ambofaciens* DSM40697, fabriquée au laboratoire à partir d'ADN total partiellement digéré par l'endonucléase *Bam*HI et cloné dans le cosmide Supercos1 (Evans *et al.*, 1989). Ces cosmides recombinants portent des inserts de taille moyenne comprise entre 30 kb et 40 kb. Le plasmide pIJ790 porte le système  $\lambda$ RED et un gène conférant la résistance au chloramphénicol. Ce système est placé sous le contrôle d'un promoteur inductible par le L-arabinose. Après induction, la souche *E. coli* BW25113/pIJ790/cosmide d'intérêt (cultivée à 30°C) est transformée par le produit de PCR (Figure R-77, cadre 5). Les transformants sont ensuite sélectionnés pour les résistances portées par le cosmide (ampicilline ou kanamycine) et par la cassette (par exemple, apramycine, résistance conférée par le gène *aac*(3)IV). La température de culture d'*E. coli* est amenée à 37°C lors de cette étape, ce qui provoque la perte du plasmide pIJ790, sa répllication étant initiée à partir d'une origine de répllication thermosensible. Après sélection des recombinants, le cosmide modifié au locus souhaité est validé et électroporé dans la souche non méthylante *E. coli* ET12567/pUZ8002. Les transformants sont sélectionnés une nouvelle fois sur la résistance portée par la cassette. Cette étape permet de contourner les éventuels systèmes de restriction-modification présents chez certaines espèces de *Streptomyces*.

### **B-IV.1.3 - TRANSFERT CONJUGATIF DU COSMIDE MODIFIÉ D'*E. COLI* VERS *STREPTOMYCES* ET SÉLECTION DES RECOMBINANTS**

Une conjugaison interspécifique, technique plus efficace que la transformation de protoplastes, est réalisée entre *E. coli* ET12567/pUZ8002/cosmide modifié et *S. ambofaciens* DSM40697 (Figure R-77, cadre 6). Le plasmide pUZ8002 porte les fonctions nécessaires à la mobilisation d'une molécule d'ADN portant une origine de transfert de type RK2 (Paget *et al.*, 1999). Le cosmide modifié ne portant pas d'origine de répllication lui permettant de se maintenir de manière autonome chez *Streptomyces*, la résistance portée par la construction ne peut être conférée que par l'intégration sur le chromosome. La recombinaison entre le cosmide modifié et le chromosome est favorisée par la longueur d'homologie entre les régions flanquant la cassette et le locus d'intérêt, longueur variant de quelques kilobases à une vingtaine de kilobases.

Les exconjugants sont d'abord sélectionnés sur la résistance portée par la cassette, puis dans un deuxième temps, testés pour leur sensibilité à la kanamycine (résistance portée par la partie cosmidiqque). De cette manière, seuls les événements de remplacement du gène d'intérêt via un événement de double crossing-over sont sélectionnés. Un crossing-over unique a pour conséquence l'intégration de la totalité du cosmide modifié dans le génome de *Streptomyces*, conférant la résistance à la kanamycine en plus de la résistance portée par la cassette.

L'intégration de la cassette de résistance au sein du chromosome étant stable, la pression de sélection peut être levée. Ceci permet notamment de comparer les phénotypes des souches mutantes et sauvage dans les mêmes conditions expérimentales.

Cette stratégie a été appliquée chez *S. ambofaciens* DSM40697 pour la fabrication de mutants des ORF dupliquées *hasL* et *hasR*. Le phénotype des mutants simples sera recherché. La construction

de mutants inactivés simultanément pour *basR* et *basL* permettra d'observer si le phénotype est plus prononcé par rapport aux mutants simples ou si en absence de phénotype chez les simples mutants, la double mutation confère un phénotype détectable.

## B-IV.2 $\propto$ CONSTRUCTION DES SIMPLES MUTANTS $\Delta basR$ ET $\Delta basL$

### B-IV.2.1 - INTRODUCTION DES COSMIDES 14C4 ET 25E1 DANS LA SOUCHE *E. COLI* BW25113/PIJ790

Les cosmides recombinants 14C4 et 25E1 renferment respectivement les loci *basL* et *basR* (voir Matériel et Méthodes) (Fischer *et al.*, 1998b). Leur insert respectif présente une taille de 42,7 kb et 39 kb. Les cosmides ont été introduits chez *E. coli* BW25113/pIJ790 (voir section Matériel et Méthodes). Treize et 40 clones ont été respectivement obtenus pour les transformations avec l'ADN des cosmides 14C4 et 25E1. L'ADN extrachromosomique de transformants cultivés à 30°C ou 37°C a été extrait et contrôlé par migration électrophorétique. Après culture à 30°C, deux bandes de haut poids moléculaire sont observées dont l'une disparaît sur le profil correspondant à l'ADN extrait après culture à 37°C. L'absence de cette bande est corrélée à la perte du plasmide pIJ790 dont la répllication est thermosensible. La bande résiduelle correspond à l'ADN natif du cosmide recombinant 14C4 ou 25E1. La digestion par *Bam*HI de l'ADN des transformants cultivés à 37°C révèle les profils attendus pour les cosmides 14C4 et 25E1 (Figure R-79).

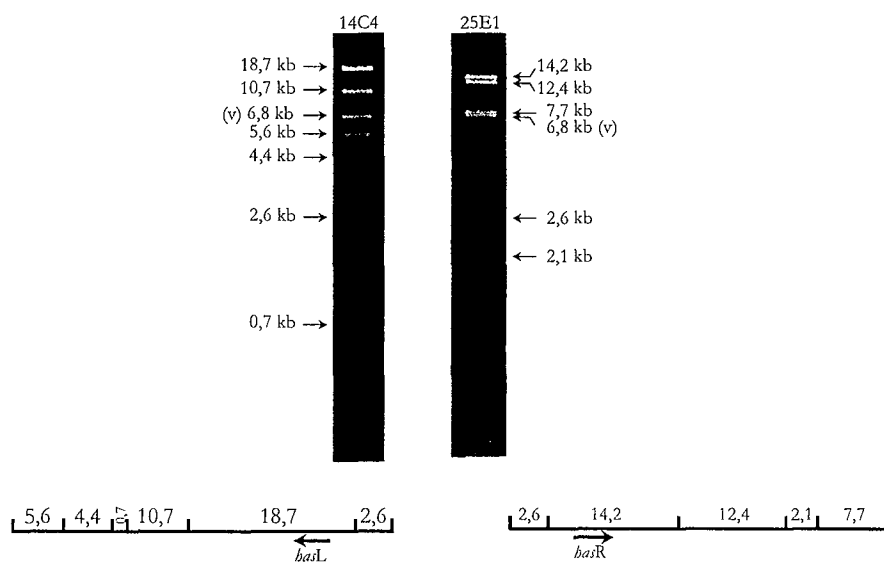


Figure R-79 : Vérification des cosmides 14C4 et 25E1 par restriction *Bam*HI.

Profils de restriction et cartes de restriction *Bam*HI des inserts de 14C4 et 25E1 et position des ORF *basR* et *basL*. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm. La taille des bandes *Bam*HI est indiquée et (v) indique le vecteur Super1 (6,8 kb).

## B-IV.2.2 – AMPLIFICATION DES CASSETTES APRA

L'amplification des cassettes est réalisée à partir de l'ADN du plasmide pIJ773 grâce aux couples d'amorces *basL1/basL2* et *basR1/basR2* (Tableau MM-6). Le produit d'amplification majoritaire a une taille estimée à 1,4 kb (Figure R-80) compatible avec la taille attendue soit 1462 pb (1384 + 2 x 39 pb). La bande quantitativement minoritaire correspond à un produit de PCR aspécifique. L'ADN amplifié est purifié à l'aide d'un kit de purification de produits de PCR (Qiagen).

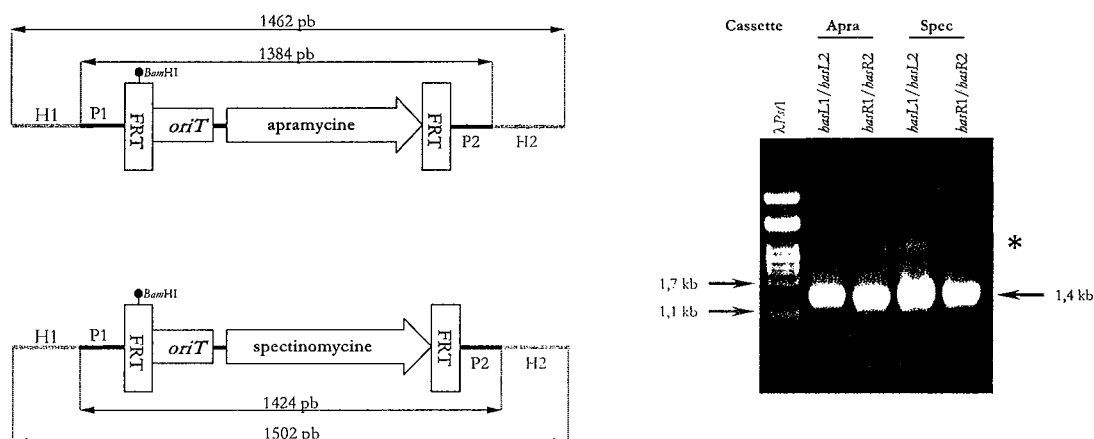


Figure R-80 : Amplification des cassettes Apra et Spec.

A gauche : Représentation des cassettes de remplacement indiquant l'emplacement du site *Bam*HI dans les cassettes.

A droite : Migration électrophorétique des produits d'amplification. *basL1/basL2* et *basR1/basR2* représentent les couples d'amorces utilisés. L'astérisque indique un produit minoritaire. Les conditions de migration sont : gel d'agarose 1%, tampon TBE, 6 V/cm. Seules les tailles des bandes du marqueur de taille  $\lambda\Psi I$  (250 ng) encadrant la bande d'intérêt sont indiquées. Un aliquot de 5  $\mu$ l sur 50  $\mu$ l du produit de PCR a été déposé.

## B-IV.2.3 – REMPLACEMENT DES ORF *HASL* ET *HASR* PAR LA CASSETTE APRA DANS LES COSMIDES RECOMBINANTS

### B-IV.2.3.1 – ELECTROPORATION ET SELECTION

Environ 100 ng de chaque amplifiat ont été utilisés pour transformer *E. coli* BW25113/pIJ790/14C4 ou 25E1. Les cellules compétentes ont été préparées à partir de cultures supplémentées en L-arabinose afin d'induire le système  $\lambda$ RED. Après transformation, les cellules sont cultivées à 37°C afin d'éliminer le plasmide pIJ790 à répllication thermosensible. Une



sélection positive basée sur la résistance à l'apramycine conférée par la cassette est effectuée. Respectivement, 110 et 9 clones [apra<sup>R</sup>] ont été dénombrés pour les transformations de BW25113/pIJ790/14C4 et 25E1. Ces clones renferment potentiellement un cosmide dont l'ORF *bas* (R ou L) a été interrompue par remplacement avec la cassette Apra.

### B-IV.2.3.2 – ANALYSE DES TRANSFORMANTS

Quatre clones issus de chaque transformation sont analysés par migration électrophorétique de leur ADN extrachromosomique digéré par *Bam*HI. Deux possibilités sont envisagées. Si le cosmide recombinant n'est pas modifié, un profil de restriction identique à celui du cosmide initial doit être observé. Dans le cas contraire, le profil du cosmide recombinant sera modifié. En effet, la cassette Apra possède un site de restriction *Bam*HI dans la séquence FRT situé en amont de l'origine de transfert *oriT* et du gène de résistance. En revanche, la séquence des ORF *bas* ne présente aucun site *Bam*HI. Le remplacement des gènes *bas* se traduit en conséquence par la disparition d'une bande au profit de deux nouvelles.

Le fragment *Bam*HI contenant le gène *basL* a une taille estimée à 18,7 kb. Le remplacement de cette ORF par la cassette Apra entraîne la disparition du fragment de 18,7 kb au profit de deux nouvelles bandes d'environ 1 kb et 18,2 kb (Figure R-81). Bien que la disparition du fragment de 18,7 kb au profit d'un fragment de 18,2 kb soit difficile à visualiser dans les conditions de migration utilisées, l'apparition du fragment de 1 kb a été utilisée comme critère pour valider les transformants à ce stade de l'analyse (Figure R-81).

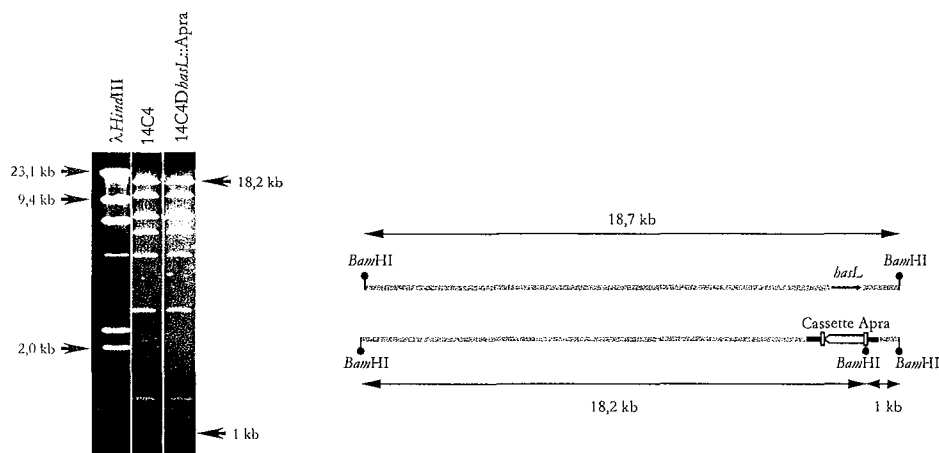
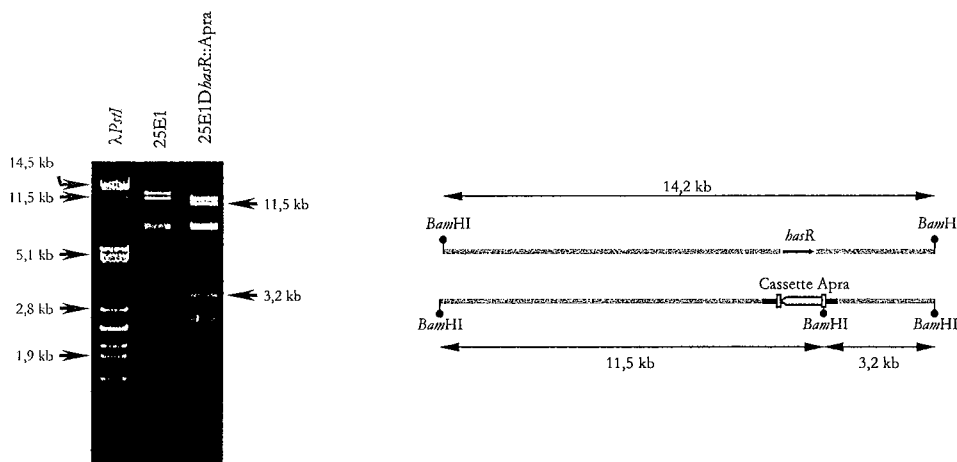


Figure R-81 : Vérification par restriction *Bam*HI du remplacement de l'ORF *hasL* par la cassette Apra.

A gauche, migration électrophorétique de l'ADN de 14C4 et 14C4 $\Delta$ *hasL*::Apra digéré par *Bam*HI. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm. Un clone représentatif sur les quatre sélectionnés est montré.

A droite, localisation de *hasL* dans le fragment *Bam*HI de 18,7 kb de 14C4 et de la cassette Apra dans les fragments de 18,2 kb et 1 kb de 14C4 $\Delta$ *hasL*::Apra.

Le changement introduit, dans le profil du cosmide 25E1, par la recombinaison de la cassette Apra entraîne le remplacement d'un fragment *Bam*HI d'environ 14,2 kb, incluant le gène *hasR*, par deux nouveaux fragments de 3,2 kb et 11,5 kb (Figure R-82). Les profils de restriction *Bam*HI obtenus pour les 4 transformants testés sont compatibles avec le remplacement du locus *hasR* par la cassette dans le cosmide 25E1 (Figure R-82).



**Figure R-82 : Vérification par restriction *Bam*HI du remplacement de l'ORF *hasR* par la cassette Apra.**

A gauche, migration électrophorétique de l'ADN de 25E1 et 25E1 $\Delta$ *hasR*::Apra digéré par *Bam*HI. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm. Un clone représentatif sur les quatre sélectionnés est montré.

A droite, localisation de *hasR* dans le fragment *Bam*HI de 14,2 kb de 25E1 et de la cassette Apra dans les fragments de 11,5 kb et 3,2 kb de 25E1 $\Delta$ *hasR*::Apra.

Les cosmides recombinants utilisés par la suite sont appelés 14C4 $\Delta$ *hasL*::Apra et 25E1 $\Delta$ *hasR*::Apra (dans l'ordre : nom du cosmide recombinant,  $\Delta$ *has* pour indiquer l'ORF remplacée et nature de la cassette utilisée pour le remplacement).

## **B-IV.2.4 - REMPLACEMENT DES ORF HAS AUX LOCI CHROMOSOMIQUES CHEZ *S. AMBOFACIENS* DSM40697**

### B-IV.2.4.1 – INTRODUCTION DES COSMIDES MODIFIES DANS LA SOUCHE *E. COLI* ET12567/PUZ8002

Afin de contourner d'éventuels systèmes de restriction-modification présents chez *S. ambofaciens*, les constructions sont introduites dans une souche d'*E. coli* déficiente dans ces mécanismes de méthylation. Ceci permet d'augmenter la probabilité d'obtenir des clones recombinants de *S. ambofaciens* lors de l'étape suivante. Une quantité d'environ 100 ng de chaque cosmide recombinant est utilisée pour transformer environ  $10^9$  cellules compétentes de la souche *E. coli*

ET12567/pUZ8002. Un crible positif est effectué sur milieu gélosé sélectif. Plusieurs centaines de clones ont été obtenus pour chaque cosmide.

B-IV.2.4.2 – CONJUGAISONS INTERGENÉRIQUES ENTRE *E. COLI*  
ET12567/pUZ8002/14C4 $\Delta$ *hasL*::APRA OU 25E1 $\Delta$ *hasR*::APRA ET  
*S. AMBOFACIENS* DSM40697

La conjugaison intergénérique est réalisée entre les souches d'*E. coli* ET12567/pUZ8002/14C4 $\Delta$ *hasL*::Apra ou 25E1 $\Delta$ *hasR*::Apra et des spores de la souche sauvage *S. ambofaciens* DSM40697. Le crible de sélection positive est la résistance des clones à l'apramycine (résistance portée par la cassette). En effet, le cosmide modifié ne portant pas d'origine de réplication lui permettant de se maintenir de manière autonome chez *Streptomyces*, la résistance ne peut être conférée que par l'intégration de la construction dans le génome (voir section Matériel et Méthodes pour les témoins). L'intégration par recombinaison est favorisée par la taille importante des régions homologues entre la construction et le génome (jusqu'à une vingtaine de kilobases de part et d'autre du locus *bas*).

Afin de distinguer les événements simples de recombinaison (un crossing-over), conduisant à l'intégration du cosmide modifié entier, des événements doubles aboutissant à l'échange allélique (remplacement du locus *bas* sauvage par la cassette), des répliques des exconjugants sur milieu contenant de la kanamycine ont été entreprises. En effet, les colonies de phénotype [apra<sup>R</sup>kan<sup>R</sup>] auront intégré le cosmide modifié entier alors que les colonies [apra<sup>R</sup>kan<sup>S</sup>] auront potentiellement échangé les allèles *bas*. L'analyse montre que tous les exconjugants [apra<sup>R</sup>] sur un effectif d'au moins 1000 clones par expériences de remplacement sont [kan<sup>S</sup>]. Ainsi, un double événement de recombinaison s'est produit chez tous les exconjugants. Ces résultats contrastent avec ceux obtenus chez *S. coelicolor* A3(2) (B. Gust, communication personnelle) et *S. ambofaciens* ATCC23877 (résultats non montrés) où la fréquence d'événements de double recombinaison avoisine généralement 20%. Ce fait indique que l'efficacité de recombinaison chez la souche *S. ambofaciens* DSM40697 semble plus importante que chez les deux autres espèces.

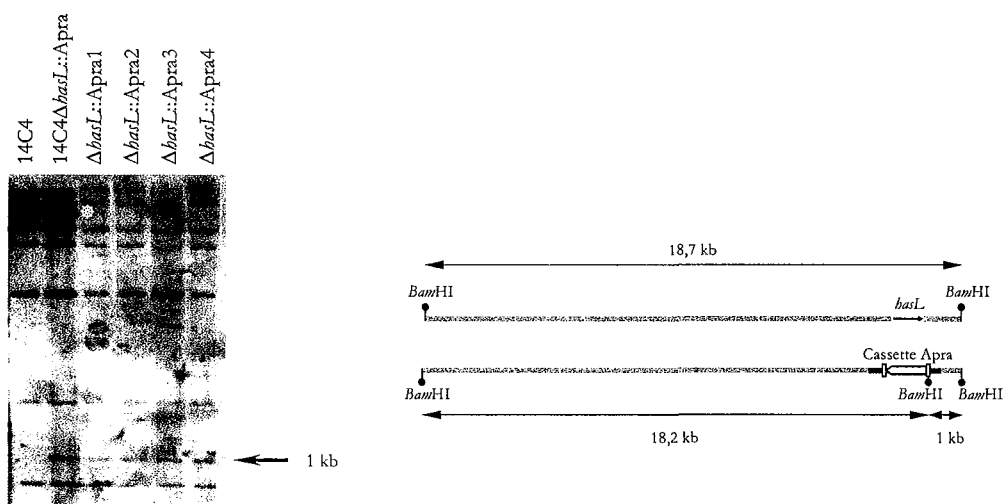
Quatre exconjugants [apra<sup>R</sup>kan<sup>S</sup>] issus de chacune des expériences de conjugaison intergénérique ont été analysés plus avant.

B-IV.2.4.3 – MISE EN EVIDENCE DE L'ÉCHANGE ALLELIQUE AUX LOCI  
*HAS* CHROMOSOMIQUES

**B-IV.2.4.3.1 – Vérification par hybridation**

Le remplacement des loci *bas* à leur locus chromosomique respectif est vérifié par hybridation d'un profil de restriction *Bam*HI de l'ADN total des quatre mutants sélectionnés par une sonde consistant en l'ADN cosmique recouvrant le locus sauvage (à savoir 14C4 pour *basL* et 25E1 pour *basR*).

L'hybridation de la sonde 14C4 sur l'ADN total des exconjugants révèle un fragment de 1 kb présent dans le cosmide modifié 14C4 $\Delta basL::Apra$  (Figure R-83). En revanche, la disparition du fragment de 18,7 kb (portant le locus *basL* sauvage) au profit d'un fragment de 18,2 kb ne peut être affirmée par la seule analyse de la seule de la Figure R-83. La disparition de ce fragment qui est compatible avec le remplacement allélique a été confirmée par la résolution des bandes de hauts poids moléculaires dans des conditions électrophorétiques adaptées (résultats non montrés). Les autres fragments révélés par la sonde sont également présents dans l'ADN cosmique 14C4 $\Delta basL::Apra$  et génomique sauvage confirmant que seul le locus *basL* a été modifié. L'absence de la bande correspondant au vecteur (Super1, 6,8 kb), dans les profils des exconjugants, confirme qu'un événement de double recombinaison s'est produit dans les quatre souches testées (en plus du fait que les souches présentent un phénotype [kan<sup>r</sup>]).

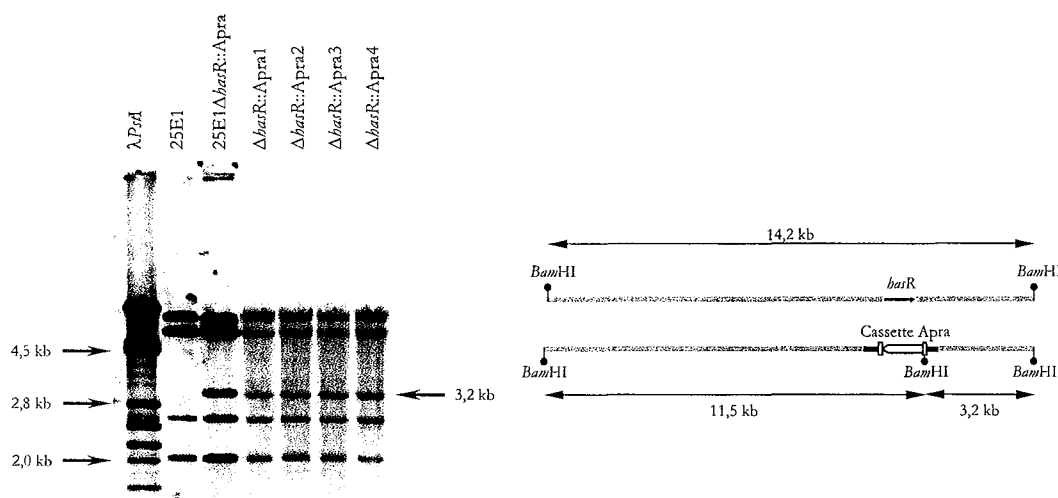


**Figure R-83 : Mise en évidence du remplacement de *basL* chromosomique par la cassette Apra.**

A gauche, hybridation de l'ADN du cosmide 14C4 marqué sur l'ADN digéré par *Bam*HI de 14C4, de 14C4 $\Delta basL::Apra$ , et de quatre mutants de *S. ambifaciens* ( $\Delta basL::Apra1$  à 4) dont l'ORF *basL* a été remplacée par la cassette Apra. L'astérisque indique le vecteur Super1 détecté dans l'ADN des cosmides. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm.

A droite, localisation de *basL* dans le fragment *Bam*HI de 18,7 kb de 14C4 et de la cassette Apra dans les fragments de 18,2 kb et 1 kb de 14C4 $\Delta basL::Apra$ .

Une démarche analogue a été appliquée pour l'analyse des quatre exconjugants correspondant au remplacement au locus *basR*. L'hybridation de la sonde 25E1 révèle chez les exconjugants une bande de 3,2 kb présente dans le profil du cosmide modifié 25E1 $\Delta basR::Apra$  (Figure R-84). Ce résultat indique que le locus *bas* a vraisemblablement été remplacé. Dans les conditions électrophorétiques utilisées, il n'est cependant pas possible de conclure quant à la disparition de la bande de 14,2 kb (portant le gène *basR*), corollaire du remplacement du locus sauvage. Il en est de même pour l'absence dans le profil des ADN génomiques de la bande caractéristique de Super1 (6,8 kb). Cette absence a toutefois été vérifiée lors d'autres expériences d'hybridation, confirmant ainsi l'occurrence du double crossing-over (résultats non montrés).



**Figure R-84 : Mise en évidence du remplacement de *hasR* chromosomique par la cassette Apra.**

A gauche, Hybridation de l'ADN du cosmide 25E1 marqué sur l'ADN digéré par BamHI de 25E1, de 25E1 $\Delta hasR::Apra$ , et de quatre mutants de *S. ambofaciens* ( $\Delta hasR::Apra1$  à 4) dont l'ORF *hasR* a été remplacé par la cassette Apra. La bande correspondant au vecteur Super1 (6,8 kb) comigre, dans les conditions de migration utilisées, avec la bande BamHI de 7,7 kb de 25E1, et n'est pas indiquée. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm.

A droite, localisation de *hasR* dans le fragment BamHI de 14,2 kb de 25E1 et de la cassette Apra dans les fragments de 11,5 kb et 1 kb de 25E1  $\Delta hasR::Apra$ .

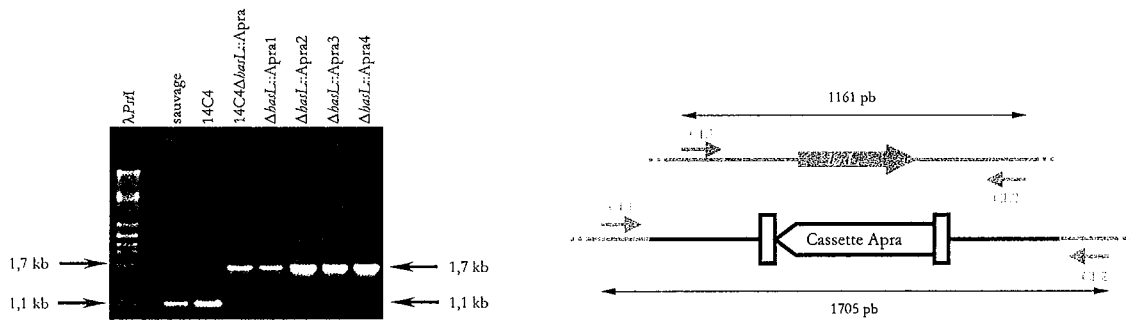
Pour confirmer le remplacement des loci *has* chez les exconjugants et donc la construction de simples mutants *has*, une analyse complémentaire a été entreprise.

#### **B-IV.2.4.3.2 – Vérification par PCR**

Pour chaque locus *has*, un couple d'amorces spécifiques des régions flanquant les deux ORF *has* a été dessiné : CL1/CL2 pour Contrôle *hasL* et CR1/CR2 pour Contrôle *hasR* (Tableau MM-6, Figures R-85 et R-86). La taille théorique des produits de PCR déduits des séquences des loci sauvages est de 1182 pb pour *hasR* et 1161 pb pour *hasL*. En revanche, si les loci sont remplacés par la cassette Apra, la taille des produits attendus passe à 1726 pb pour un simple mutant *S. ambofaciens*  $\Delta hasR::Apra$  et 1705 pb pour un simple mutant *S. ambofaciens*  $\Delta hasL::Apra$ .

L'amplification à l'aide du couple CL1/CL2 (Figure R-85) permet d'obtenir une bande unique de taille estimée à 1,1 kb lorsque l'ADN génomique sauvage et l'ADN du cosmide 14C4 sont utilisés comme matrice, ce qui est compatible avec la taille attendue de 1162 pb. En revanche, lorsque l'ADN génomique des quatre exconjugants et l'ADN du cosmide 14C4 $\Delta hasL::Apra$  sont utilisés en tant que matrice, une bande unique d'environ 1,7 kb est obtenue, ce qui est compatible avec la

taille théorique attendue. Ces données démontrent le remplacement effectif de *basL* par l'allèle portant la cassette Apra au locus chromosomique.

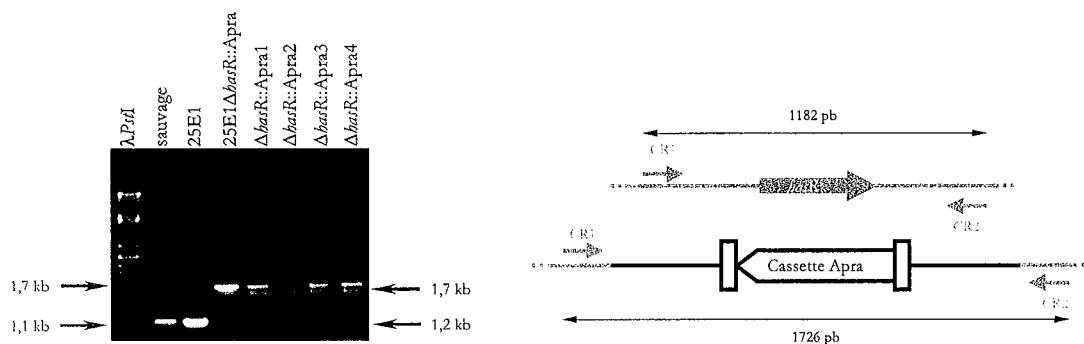


**Figure R-85 : Vérification par PCR du remplacement de l'ORF *hasL*.**

A gauche, séparation des produits de PCR sur un gel d'agarose en fonction des matrices : ADN de la souche sauvage de *S. ambifaciens* DSM40697, de 14C4, de 14C4 $\Delta$ *hasL*::Apra, et de quatre mutants de la souche *S. ambifaciens* DSM40697 ( $\Delta$ *hasL*::Apra1 à 4) portant l'allèle *hasL* remplacé par la cassette Apra. Les conditions de migration sont : gel d'agarose 1%, tampon TBE, 6 V/cm. Seules les tailles des bandes du marqueur de taille  $\lambda$ *PstI* (250 ng) encadrant les bandes d'intérêt sont indiquées. Un aliquot de 5  $\mu$ l sur 50  $\mu$ l du produit de PCR a été déposé.

À droite, représentation schématique de l'organisation des loci *hasL* sauvage (partie haute) et remplacé par la cassette Apra (partie basse). Les oligonucléotides (CL1 et CL2) servant à la PCR sont représentés par des flèches et les tailles des produits de PCR attendus sont indiquées.

Des résultats analogues ont été obtenus pour le locus *basR* (Figure R-86) et permettent de conclure au remplacement efficace du locus *basR* dans les quatre exconjugants sélectionnés.



**Figure R-86 : Vérification par PCR du remplacement de l'ORF *hasR*.**

A gauche, séparation des produits de PCR sur un gel d'agarose en fonction des matrices : ADN de la souche sauvage de *S. ambifaciens* DSM40697, de 25E1, de 25E1 $\Delta$ *hasR*::Apra, et de quatre mutants de la souche *S. ambifaciens* DSM40697 ( $\Delta$ *hasR*::Apra1 à 4) portant l'allèle *hasR* remplacé par la cassette Apra. Les conditions de migration sont : gel d'agarose 1%, tampon TBE, 6 V/cm. Seules les tailles des bandes du marqueur de taille  $\lambda$ *PstI* (250 ng) encadrant les bandes d'intérêt sont indiquées. Un aliquot de 5  $\mu$ l sur 50  $\mu$ l du produit de PCR a été déposé.

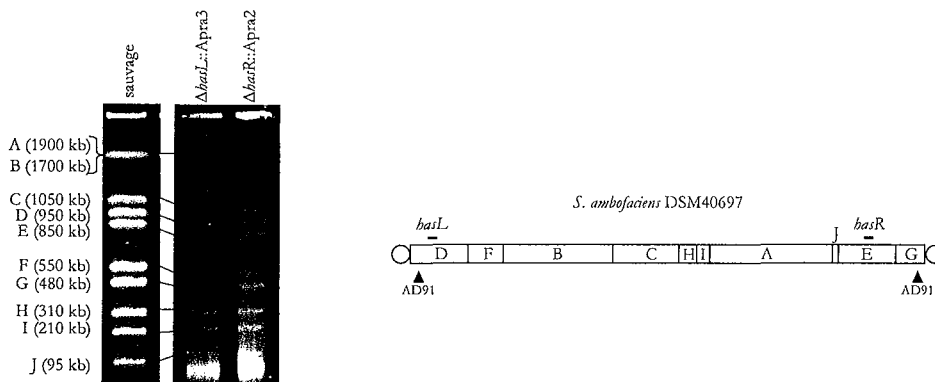
À droite, représentation schématique de l'organisation des loci *hasR* sauvage (partie haute) et remplacé par la cassette Apra (partie basse). Les oligonucléotides (CR1 et CR2) servant à la PCR sont représentés par des flèches et les tailles des produits de PCR attendus sont indiquées.

B-IV.2.4.4 – VÉRIFICATION DE L'INTEGRITÉ DE LA STRUCTURE  
CHROMOSOMIQUE CHEZ LES SIMPLES MUTANTS *S. AMBOFACIENS*  
 $\Delta HAS::APRA$

Le chromosome de *S. ambofaciens* DSM40697 est sujet à une forte instabilité structurale notamment dans les régions subtélomériques (pour revue, (Leblond and Decaris, 1994). Les loci *bas* étant localisés dans les régions terminales, il était nécessaire de vérifier qu'aucun réarrangement de grande ampleur n'avait accompagné le remplacement des allèles *bas*.

Pour cela, deux approches ont été utilisées : l'analyse de la structure génomique par électrophorèse en champs pulsés (PFGE pour "Pulsed Field Gel Electrophoresis"), et l'hybridation de contrôle par l'ADN du cosmide AD91 recouvrant les régions terminales du chromosome (Leblond *et al.*, 1996).

La comparaison en PFGE des profils *AseI* de l'ADN du sauvage avec celui des simples mutants ne révèle aucune différence décelable (Figure R-87), notamment pour les fragments portant les loci *bas*, fragment *AseI*-D (950 kb) pour *basL*, et fragment *AseI*-E (850 kb) pour *basR*. Ainsi, aucun réarrangement de grande ampleur ne s'est produit lors de la construction des simples mutants *S. ambofaciens*  $\Delta bas::Apra$ .



**Figure R-87 : Vérification de l'intégrité du chromosome des simples mutants.**

A gauche, migration de digestions *AseI* des ADN sauvage et de deux simples mutants ( $\Delta basL::Apra3$  et  $\Delta basR::Apra2$ ) en PFGE. Le nom et la taille de chaque fragment *AseI* sont indiqués. Les conditions de migration n'étant pas les mêmes entre la piste de l'ADN sauvage (gel d'agarose 0,9%, 6V/cm, tampon TBE, temps de pulsation 40-160 s, 20 h) et celles des mutants (paramètres identiques sauf la durée de migration ramenée à 24 h), les correspondances entre les bandes sont indiquées par des traits. A droite, carte de restriction *AseI* du génome de *S. ambofaciens* DSM40697. Les rectangles correspondant aux fragments de restriction *AseI* sont désignés par une lettre (A à J). La localisation des loci *basL* et *basR* dans les fragments D et E est signalée par un trait horizontal. Les deux cercles symbolisent les protéines terminales liées de façon covalente à l'ADN. Les triangles noirs indiquent la localisation du cosmide AD91.

De plus, lorsque les ADN génomiques des simples mutants  $\Delta bas::Apra$  sont été hybridés avec l'ADN marqué du cosmide recombinant AD91, un profil identique à celui de la souche sauvage a

été obtenu pour les 8 souches testées indiquant que les extrémités chromosomiques sont vraisemblablement présentes et intactes (résultats non montrés).

Deux simples mutants ont été analysés par la suite et seront appelés  $\Delta hasL::Apra$  et  $\Delta hasR::Apra$  ( $\Delta hasL$  ou  $\Delta hasR$  désigne l'ORF remplacée, respectivement *hasL* ou *hasR* ;  $::Apra$  indique la nature de la cassette de remplacement).

### **B-IV.3 $\propto$ CONSTRUCTION DE DOUBLES MUTANTS $\Delta HASR \Delta HASL$**

La quasi identité (97%) des protéines potentiellement codées par les loci sauvages *has* permettait d'envisager une redondance fonctionnelle des deux allèles. C'est pourquoi l'interruption des deux copies au sein du même génome a été entreprise. L'approche utilisée pour interrompre la copie de l'ORF *has* résiduelle chez les simples mutants  $\Delta hasL::Apra$  et  $\Delta hasR::Apra$  est analogue à celle employée pour interrompre l'une des deux copies chez la souche sauvage. Le fait d'interrompre l'ORF *hasL* chez le simple mutant  $\Delta hasR::Apra$  et réciproquement permet d'obtenir des doubles mutants indépendants, et assure ainsi la validité des analyses lors de la recherche de phénotypes.

#### **B-IV.3.1 – AMPLIFICATION DES CASSETTES SPEC ET REMPLACEMENT DIRIGE DES LOCI *HAS* SUR LES COSMIDES RECOMBINANTS 14C4 ET 25E1**

Les séquences bordant les différentes cassettes étant identiques, les couples d'amorces *hasL1/hasL2* et *hasR1/hasR2* (Tableau MM-6) ont été utilisés pour amplifier par PCR la cassette Spec (gène *aadA* et *oriT*) à partir de l'ADN du plasmide pIJ778 (voir Matériel et Méthodes). Une bande d'environ 1,5 kb est observée sur la Figure R-80 et correspond à la taille attendue de l'amplifiat (1502 pb). Une bande quantitativement minoritaire de taille plus importante est également observée. L'ADN amplifié est purifié à l'aide d'un kit de purification (Qiagen).

Des transformants d'*E. coli* BW25113/pIJ790/14C4 et 25E1 sont alors obtenus suite à l'électroporation avec les produits de PCR respectifs (respectivement 227 et 11 clones [*spec*<sup>R</sup>]). Trois clones [*spec*<sup>R</sup>] sont sélectionnés parmi les transformants obtenus pour chaque expérience. L'analyse par restriction *Bam*HI de l'ADN cosmétique des 3 clones sélectionnés chez la souche BW25113/pIJ790/14C4 (Figure R-88) révèle une bande de 1 kb correspondant au remplacement de *hasL* et à l'obtention de la construction 14C4 $\Delta hasL::Spec$ .



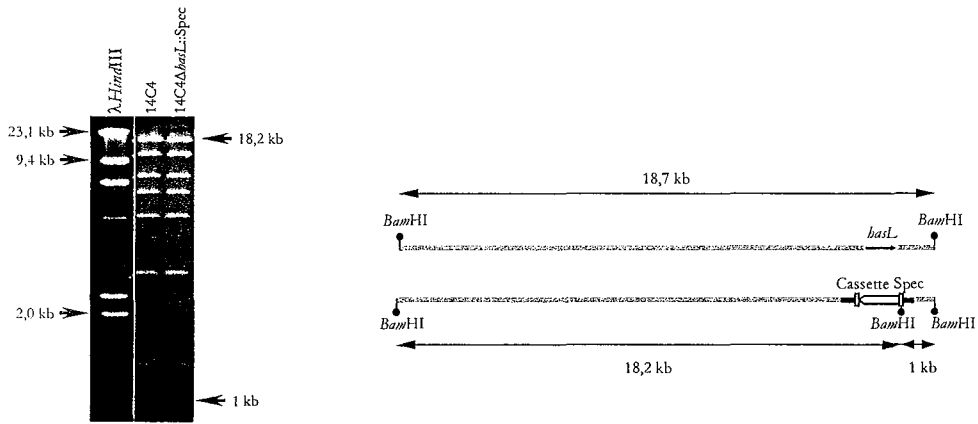


Figure R-88 : Vérification par restriction *Bam*HI des remplacements de l'ORF *hasL* par la cassette Spec.

A gauche, migration électrophorétique de l'ADN des cosmides 14C4 et 14C4 $\Delta$ *hasL*::Spec digéré par *Bam*HI. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm. Un clone représentatif sur les quatre sélectionnés est montré.

A droite, localisation de *hasL* dans le fragment *Bam*HI de 18,7 kb de 14C4 et de la cassette Spec dans les fragments de 18,2 kb et 1 kb de 14C4 $\Delta$ *hasL*::Spec.

De même, l'observation d'une bande d'environ 3,2 kb dans le profil de restriction *Bam*HI des ADN cosmidiqes des 3 clones sélectionnés à partir de la souche BW25113/pIJ790/25E1 révèle le remplacement de *hasR* (Figure R-89) et l'obtention de la construction 25E1 $\Delta$ *hasR*::Spec.

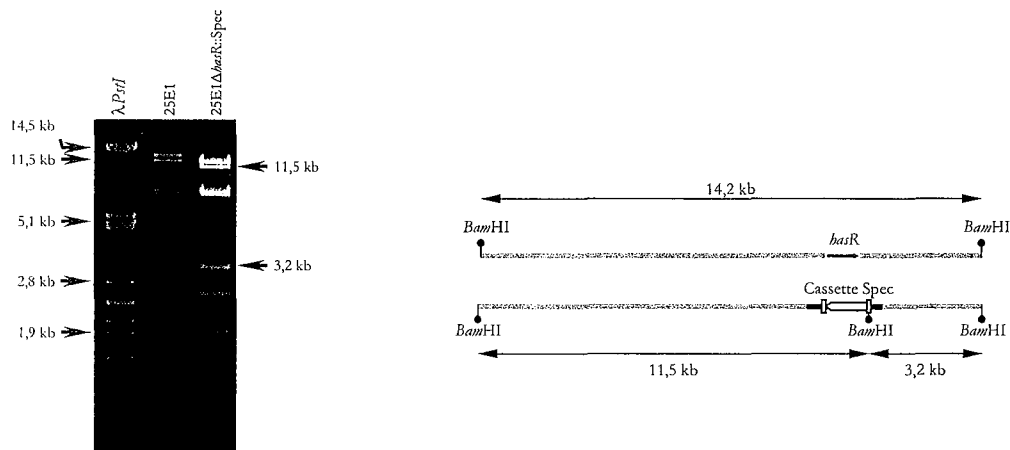


Figure R-89 : Vérification par restriction *Bam*HI du remplacement de l'ORF *hasR* par la cassette Spec.

A gauche, migration électrophorétique de l'ADN de 25E1 et 25E1 $\Delta$ *hasR*::Spec digéré par *Bam*HI. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm. Un clone représentatif sur les quatre sélectionnés est montré.

A droite, localisation de *hasR* dans le fragment *Bam*HI de 14,2 kb de 25E1 et de la cassette Spec dans les fragments de 11,5 kb et 3,2 kb de 25E1 $\Delta$ *hasR*::Spec.

Les cosmides recombinants 14C4 $\Delta hasL$ ::Spec et 25E1 $\Delta hasR$ ::Spec analysés ont été introduits dans la souche *E. coli* ET12567/pUZ8002. Parmi les transformants [spec<sup>R</sup>] obtenus (plusieurs centaines pour chaque expérience), deux clones ont été analysés et la présence des cosmides modifiés a été vérifiée par analyse de restriction *Bam*HI de l'ADN cosmétique.

### B-IV.3.2 - REMPLACEMENT DU LOCUS *HAS* SAUVAGE PAR LA CASSETTE SPEC CHEZ LES SIMPLES MUTANTS $\Delta HASL$ ::APRA ET $\Delta HASR$ ::APRA

Le cosmide 14C4 $\Delta hasL$ ::Spec a été transféré par conjugaison intergénérique chez un simple mutant  $\Delta hasR$ ::Apra (84 clones obtenus). Réciproquement, 25E1 $\Delta hasR$ ::Spec est transféré dans le simple mutant  $\Delta hasL$ ::Apra (277 clones obtenus). La totalité des exconjugants obtenus sont [kan<sup>S</sup>]. Trois clones [apra<sup>R</sup>spec<sup>R</sup>kan<sup>S</sup>] issus de chaque expérience ont été sélectionnés et analysés. Les doubles mutants sont appelés DMRL (Double Mutant dont la copie *hasL* a été remplacée par la cassette Spec dans un contexte  $\Delta hasR$ ::Apra) et DMLR (Double Mutant dont la copie *hasR* a été remplacée par la cassette Spec dans un contexte  $\Delta hasL$ ::Apra).

### B-IV.3.3 - MISE EN EVIDENCE DU REMPLACEMENT DES DEUX ORF *HAS* AUX LOCI CHROMOSOMIQUES

#### B-IV.3.3.1 - VERIFICATION PAR HYBRIDATION

L'hybridation de l'ADN marqué des cosmides 14C4 et 25E1 sur les ADN des doubles mutants révèle les bandes *Bam*HI signatures du remplacement des loci respectifs dans les contextes simples mutants (1 kb pour l'interruption de *hasL* chez un mutant *hasR*, et 3,2 kb pour l'interruption réciproque) (Figure R-90).

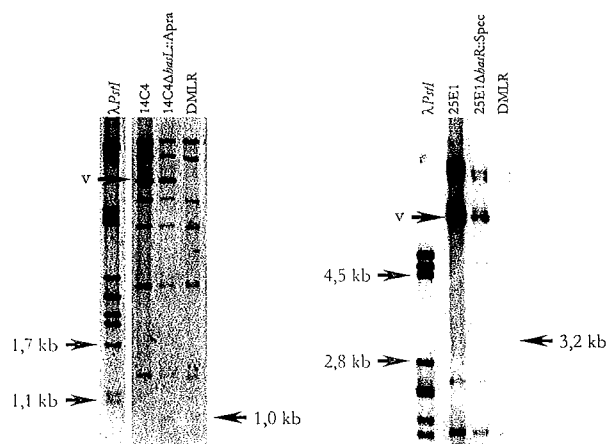


Figure R-90 : Vérification du remplacement des ORF *hasL* et *hasR* par les cassettes Apra et Spec respectivement.

A gauche, hybridation de l'ADN marqué du cosmide 14C4 sur des profils de restriction *Bam*HI des cosmides 14C4 et 14C4 $\Delta hasL$ ::Apra et de l'ADN du double mutant DMLR2.

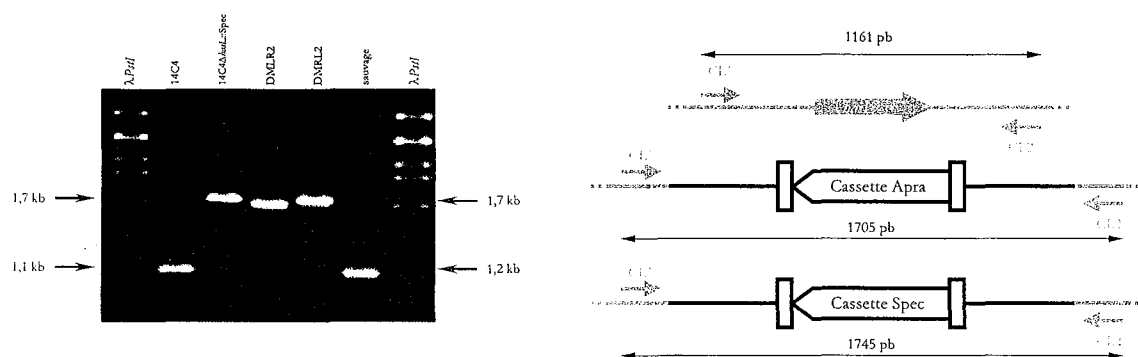
A droite; hybridation de l'ADN marqué du cosmide 25E1 sur des profils de restriction *Bam*HI des cosmides 25E1 et 25E1 $\Delta hasR$ ::Spec et de l'ADN de DMLR2.

La bande correspondant au vecteur Super1 est indiquée (v).

B-IV.3.3.2 – VERIFICATION PAR PCR

Les couples d'amorces CL1/CL2 et CR1/CR2 sont utilisés dans des expériences de PCR avec pour matrice l'ADN extrait des clones précédemment analysés ainsi que l'ADN des contrôles (14C4, 14C4 $\Delta hasL$ ::Spec, 25E1, 25E1 $\Delta hasR$ ::Spec et *S. ambifaciens* DSM40697). La taille des fragments de PCR attendus et l'analyse de deux doubles mutants réciproques sont présentés en Figure R-91.

Ainsi, l'amplification à l'aide du couple d'amorces CL1/CL2 avec pour matrice l'ADN du double mutant DMLR génère une bande unique compatible avec la taille théorique (1705 pb), confirmant le remplacement du gène *hasL* par la cassette Apra. Cette taille est légèrement inférieure à celle du produit obtenu pour l'amplification du cosmide 14C4 $\Delta hasL$ ::Spec, la cassette Spec mesurant 1745 pb.



**Figure R-91 : Vérification par PCR du remplacement de l'ORF *hasL* chez les doubles mutants.**

A gauche, séparation des produits de PCR obtenus à l'aide des oligonucléotides CL1 et CL2 en fonction de la matrice utilisée (ADN des cosmides 14C4 et 14C4 $\Delta hasL$ ::Spec, des doubles mutants DMLR2 et DMRL2 ainsi que de la souche sauvage). Les conditions de migration sont : gel d'agarose 1%, tampon TBE, 6 V/cm. Seules les tailles des bandes du marqueur de taille  $\lambda PstI$  (250 ng) encadrant les bandes d'intérêt sont indiquées. Un aliquot de 5  $\mu$ l sur 50  $\mu$ l du produit de PCR a été déposé.

A droite, représentation schématique de l'organisation des loci *hasL* sauvage (en haut) et remplacés par les cassettes Apra (au milieu) et Spec (en bas). Les oligonucléotides (CL1 et CL2) servant à la PCR sont représentés par des flèches et les tailles des produits de PCR attendus sont indiquées.

De même, l'amplification à l'aide du couple d'amorces CR1/CR2 avec pour matrice l'ADN du double mutant DMLR et du cosmide modifié 25E1 $\Delta hasR$ ::Spec génère un produit unique, identique dans les deux expériences, et de taille compatible avec le remplacement du gène *hasR* par la cassette Spec (Figure R-92).

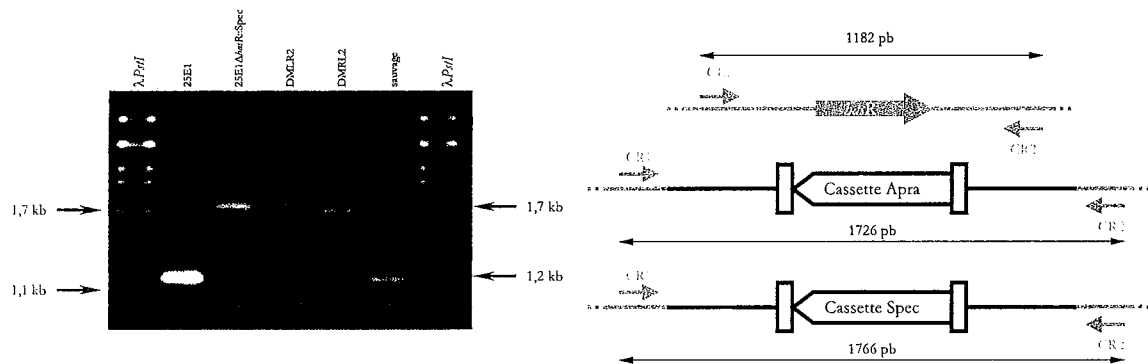


Figure R-92 : Vérification par PCR du remplacement de l'ORF *hasR* chez les doubles mutants.

A gauche, séparation des produits de PCR obtenus à l'aide des oligonucléotides CR1 et CR2 en fonction de la matrice utilisée (ADN des cosmides 25E1 et 25E1 $\Delta hasR::Spec$ , des doubles mutants DMLR2 et DMRL2 ainsi que de la souche sauvage). Les conditions de migration sont : gel d'agarose 1%, tampon TBE, 6 V/cm. Seules les tailles des bandes du marqueur de taille  $\lambda PstI$  (250 ng) encadrant les bandes d'intérêt sont indiquées.

A droite, représentation schématique de l'organisation des loci *hasR* sauvage (en haut) et remplacés par les cassettes Apra (au milieu) et Spec (en bas). Les oligonucléotides (CR1 et CR2) servant à la PCR sont représentés par des flèches et les tailles des produits de PCR attendus sont indiquées.

Réciproquement, la comparaison entre la taille des produits de PCR obtenus à partir de l'ADN du double mutant DMRL et de l'ADN des cosmides 14C4 $\Delta hasL::Spec$  et 25E1 $\Delta hasR::Spec$ , confirme à la fois le remplacement du gène *hasR* par la cassette Apra et celui du gène *hasL* par la cassette Spec.

Ces résultats confirment donc le remplacement en deux étapes des deux ORF *has* aux loci chromosomiques, et par conséquent l'obtention de doubles mutants réciproques et indépendants.

#### IV.3.4 - INTEGRITE DES REGIONS TERMINALES DU CHROMOSOME DES DOUBLES MUTANTS

De façon analogue à l'analyse du génome des simples mutants, l'intégrité structurale du chromosome des doubles mutants a été testée par PFGE et par hybridation avec le cosmide AD91 recouvrant les régions subtélomériques du chromosome sauvage (Leblond *et al.*, 1996).

Comme le montre la Figure R-93, les profils *AseI* de l'ADN de la souche sauvage et doubles mutants sont identiques.

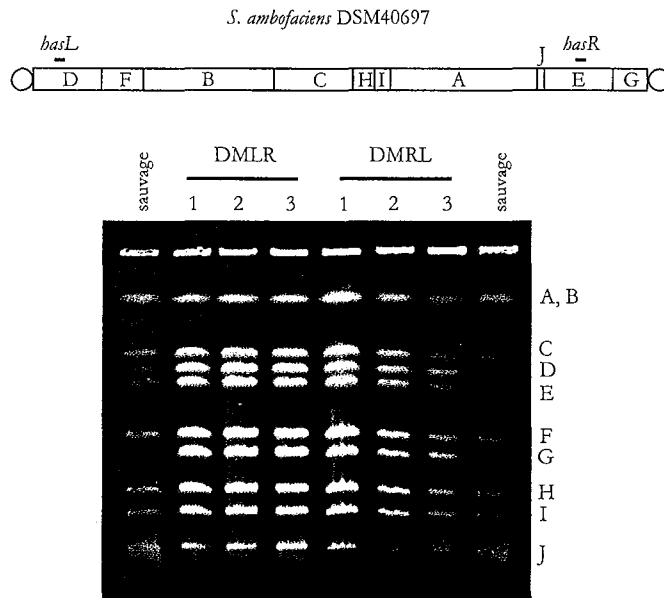


Figure R-93 : Vérification de l'intégrité du chromosome des doubles mutants.

En haut, carte de restriction *AseI* du génome de *S. ambifaciens* DSM40697. Les rectangles correspondant aux fragments de restriction *AseI* sont désignés par une lettre (A à J). La localisation des loci *basL* et *basR* dans les fragments D et E est signalée par un trait horizontal. Les deux cercles font référence aux protéines terminales liées de façon covalente à l'ADN.

En bas, migration en PFGE des ADN des doubles mutants. Le nom des fragments *AseI* est indiqué à droite du gel. Les conditions de migration sont : gel d'agarose 0,9%, 6 V/cm, tampon TBE, temps de pulsation 40-160 s, 20 h.

De même, les profils d'hybridation des ADN totaux sauvage et doubles mutants par la sonde AD91 ne présentent aucune différence (Figure R-94).

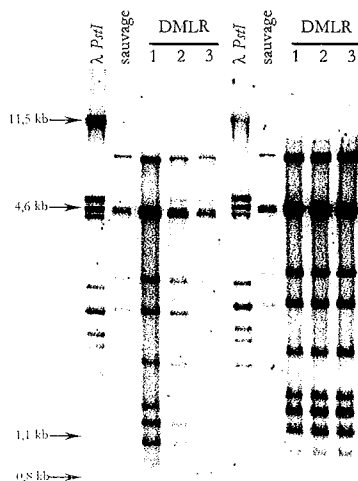


Figure R-94 : Vérification de l'intégrité des extrémités du chromosome chez les doubles mutants (DMLR et DMRL).

Hybridation de l'ADN marqué du cosmide AD91 sur des profils de restriction *BamHI* de l'ADN total de la souche sauvage et de 6 souches doubles mutantes. Les conditions de migration sont : gel d'agarose 0,8%, tampon TAE, 1,5 V/cm. Seules quelques tailles de bandes du marqueur de taille  $\lambda PstI$  (250 ng) sont indiquées.

Cette expérience révèle qu'au moins une extrémité chromosomique intacte est présente chez les souches mutantes. Comme les deux fragments *AseI* terminaux sont présents dans les profils PFGE, il est possible de conclure que les régions terminales du chromosome sont intactes dans les souches doubles mutantes.

Par conséquent, huit simples mutants indépendants issus de la souche *S. ambofaciens* DSM40697 inactivés pour *basL* et *basR* (quatre de chaque) ainsi que six doubles mutants indépendants (par groupe de trois réciproques) présentant les deux copies *bas* inactivées sont désormais disponibles pour l'analyse phénotypique.

## B-IV.4 $\otimes$ CONSTRUCTION DE COMPLÉMENTANTS DES SIMPLES MUTANTS $\Delta bas$

### B-IV.4.1 – STRATEGIE

Afin de vérifier que le phénotype des mutants interrompus dans le gène *basR* ou *basL* est effectivement dû à l'absence de gène(s) *bas* fonctionnel(s), une approche de complémentation a été envisagée. Cette approche consiste à intégrer à un autre locus chromosomique une copie sauvage du gène *bas* considéré.

Dans cette optique, les régions promotrices et codantes des gènes *basR* et *basL* ont été clonées dans le plasmide conjugatif et intégratif pSET $\Omega$  (O'Connor *et al.*, 2002). Ce plasmide a été construit à partir du plasmide conjugatif et intégratif pSET152, lui-même dérivé du phage  $\Phi$ C31 (Bierman *et al.*, 1992), où le gène *aac(3)IV* conférant la résistance à l'apramycine a été remplacé par le gène *aadA* codant la résistance aux antibiotiques spectinomycine / streptomycine. Outre le site d'intégration *attP*, ce dérivé possède l'intégrase du phage et non son excisionase ce qui lui permet de s'intégrer de façon stable au site unique d'intégration du bactériophage  $\Phi$ C31 (Kuhstoss and Rao, 1991), localisé dans le fragment *AseI*-C du chromosome de *S. ambofaciens* (Leblond *et al.*, 1996).

Il possède également une origine de répllication *colEI* qui lui permet de se répliquer chez *Escherichia coli*, ainsi que l'origine de transfert du plasmide RK2, qui lui permet d'être transféré d'*E. coli* à *Streptomyces* au cours d'une conjugaison intergénérique.

Enfin, il possède le gène de résistance aux antibiotiques spectinomycine / streptomycine (gène *aadA*), ce qui permet de sélectionner la présence du plasmide à la fois chez *E. coli* et chez *Streptomyces*.

### B-IV.4.2 – CLONAGE DES REGIONS CODANTES DES GENES HASL ET HASR DANS LE VECTEUR PSET $\Omega$

Les gènes *has* ont été isolés par PCR à l'aide des couples d'oligonucléotides CoR1/CoR2 et CoL1/CoL2 (Tableau MM-6) permettant d'amplifier des séquences de 1076 et 1176 pb correspondant à la région codante des gènes *hasR* et *hasL* respectivement, ainsi qu'à leur région promotrice.

Ces produits de PCR ont ensuite été digérés par l'endonucléase *EcoRI*, présentant un site de restriction dans chacune des deux amorces. Parallèlement, l'ADN du plasmide pSET $\Omega$  a été extrait, digéré par *EcoRI* puis déphosphorylé. Après plusieurs étapes de purification, deux mélanges de ligation ont été réalisés avec une multiplicité d'environ une molécule d'ADN plasmidique pour dix molécules d'ADN de chaque produit de PCR puis vérifiée par électrophorèse en gel d'agarose.

L'ADN de ces mélanges de ligation ainsi que l'ADN natif du vecteur pSET $\Omega$  ont enfin été électroporés dans des cellules compétentes d'*E. coli* DH5 $\alpha$ . Le vecteur de clonage pSET $\Omega$  renferme également le gène *lacZ* permettant la sélection des clones transformés par un vecteur recombinant (clones blancs sur milieu LB+X-Galactose+IPTG) dans le cas des mélanges de ligation pSET $\Omega$  + CoR et pSET $\Omega$  + CoL. Ainsi, 12 et 5 clones blancs, issus de la transformation avec les mélanges de ligation renfermant *hasR* et *hasL* respectivement, ont été analysés par migration sur gel d'électrophorèse de leur ADN plasmidique digéré par *EcoRI*.

Deux types de profils de restriction ont été observés pour chaque clonage : des profils ne présentant qu'une bande unique correspondant au plasmide vide et des profils de vecteurs recombinants où deux bandes sont observées: une bande aux alentours de 1 kb (exactement 1076 pb pour *hasR* et 1176 pb pour *hasL*) correspondant à la digestion par *KpnI* du fragment amplifié par PCR et la bande signature du vecteur vide d'environ 7 kb (exactement 6,7 kb).

Les plasmides ayant intégré le gène *hasR* ou *hasL* ont été appelés pSET $\Omega$ R et pSET $\Omega$ L respectivement.

L'insert de ces plasmides a ensuite été séquencé afin de vérifier qu'aucune mutation n'avait été introduite, notamment lors des étapes de PCR.

### B-IV.4.3 – INTRODUCTION CHEZ *STREPTOMYCES*

Des cellules compétentes d'*E. coli* ET12567 + pUZ8002 ont ensuite été transformées par l'ADN d'un plasmide recombinant de chaque type (pSET $\Omega$ R1 et pSET $\Omega$ L1) ainsi que par de l'ADN natif du vecteur pSET $\Omega$ .

L'ADN plasmidique des transformants obtenus a ensuite été extrait et des digestions *EcoRI* ont permis de vérifier la carte de restriction des vecteurs recombinants (Figure R-95).

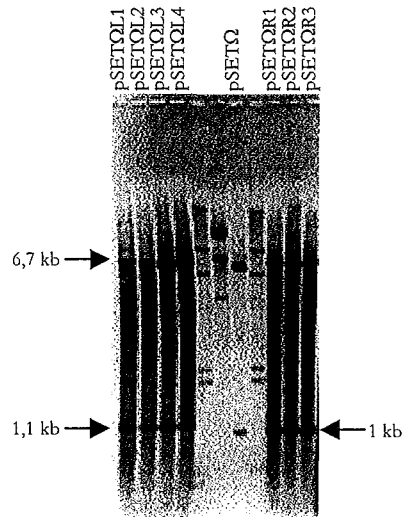


Figure R-95 : Vérification des cartes de restriction *EcoRI* des plasmides pSETΩR et pSETΩL.

Après avoir vérifié la capacité de transfert et d'intégration des plasmides recombinants par des expériences de conjugaison intraspécifique *E. coli* / *E. coli*, un plasmide recombinant représentatif de chaque construction a été introduit chez les simples mutants *has* par conjugaison intergénérique.

Après sélection sur la résistance à la spectinomycine, deux types de transconjugants issus du simple mutants R2 ont été obtenus : des transconjugants ayant intégré un vecteur vide et des transconjugants ayant intégré un vecteur recombinant pSETΩR renfermant une copie sauvage du gène *hasR*.

Le phénotype des transconjugants obtenus est de type sauvage c'est-à-dire que les colonies présentent un mycélium aérien (phénotype amy<sup>+</sup>) et une pigmentation grise (phénotype pig<sup>+</sup>). La différenciation de *S. ambifaciens* ne semble donc pas affectée par l'intégration du plasmide dans le chromosome bactérien.

Les souches de *S. ambifaciens* ainsi transformées ont été nommées NSApSETΩ et NSApSETΩR (NSA pour Nancy S. *ambifaciens*).

De façon analogue aux vérifications réalisées pour les simples et doubles mutants, l'intégrité structurale du chromosome des complémentants a été testée par PFGE et par hybridation avec le cosmide AD91. L'intégration des plasmides pSETΩ et pSETΩR a été vérifiée par des hybridations moléculaires, à l'aide de sondes correspondant au locus *has* et au site *att* chromosomique et également par PCR (résultats non montrés).



- CHAPITRE B-V -

ANALYSE PHENOTYPIQUE DES  
MUTANTS  $\Delta HAS$

Les facteurs  $\sigma$  alternatifs sont fréquemment impliqués dans la régulation de la différenciation morphologique des bactéries du genre *Streptomyces* et/ou dans le contrôle de régulateurs de réponse au stress (voir Introduction). Par ailleurs, en conditions de laboratoire, la différenciation morphologique et métabolique des *Streptomyces* est étroitement dépendante de la composition du milieu de croissance. Ainsi, selon le milieu, le phénotype colonial peut présenter une différenciation incomplète, ou au contraire, une surproduction de composés associés au métabolisme secondaire tels que les antibiotiques. La recherche de phénotypes associés à l'interruption du ou des gènes *has* est basée, d'une part sur la culture de simples et doubles mutants sur des milieux de composition variée (milieux riches et milieux minimum) et d'autre part, sur la confrontation avec divers agents stressants.

## B-V.1 $\sigma^{Has}$ LES FACTEURS $\sigma^{Has}$ ET LA DIFFERENCIATION CELLULAIRE

### B-V.1.1 – CROISSANCE SUR MILIEUX GELOSES

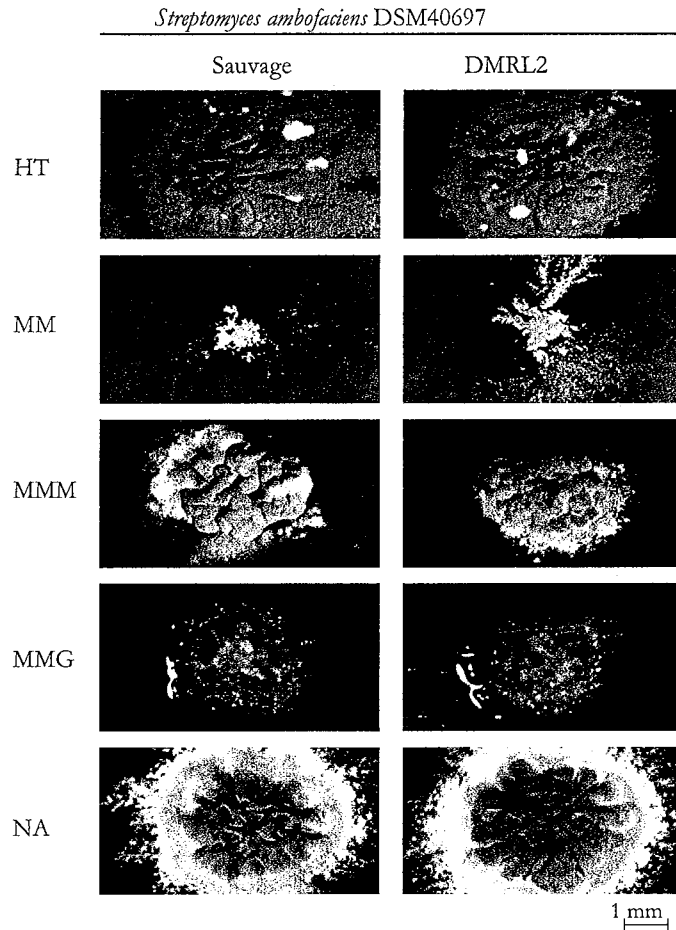
Pour tous les milieux testés, plus d'une centaine de colonies de chaque souche a été analysée.

Sur milieu R2YE (Thompson *et al.*, 1980), le phénotype des colonies mutées dans le gène *sigB* codant un facteur  $\sigma$  alternatif chez *S. coelicolor* A3(2), est déficient pour la production de mycélium aérien (phénotype "bald") (Cho *et al.*, 2001). Les ORF *has* de *S. ambofaciens* étant fortement homologues au gène *sigB* de *S. coelicolor* A3(2), les simples et doubles mutants interrompus pour la ou les ORF *has* ont été cultivés sur milieu R2YE.

Aucune différence morphologique ou retard de croissance entre les mutants et la souche sauvage n'est observée (résultat non montré). Les colonies présentent après quatre jours de croissance à 30°C une différenciation de type sauvage (formation du mycélium aérien ou [amy<sup>+</sup>], sporulation normale ou [spo<sup>+</sup>] et pigmentation grise des spores ou [pig<sup>+</sup>]).

Sur milieu R2 (Okanishi *et al.*, 1974) correspondant au milieu R2YE sans extrait de levure (YE pour "Yeast Extract"), un mutant de *S. ambofaciens* ATCC23877 interrompu pour une ORF codant une polycétone synthétase est affecté dans la production d'un pigment brun-orange diffusible (Pang *et al.*, sous presse). Sur ce même milieu, les doubles mutants *has* présentent un phénotype sauvage.

D'autres milieux riches tels que HT (Pridham *et al.*, 1957b), SFM (Kieser *et al.*, 2000), NA et milieux minimums tels que MM (Hopwood, 1967) MM complété en glucose (10 g/l) ou mannitol (5 g/l) ont été testés. Sur tous ces milieux, les simples et doubles mutants *has* ne présentent pas de différence phénotypique vis-à-vis du sauvage (Figure R-96). Hormis sur MM glucose où les colonies sauvages et mutantes arborent une sporulation pauvre, les colonies ont un phénotype [amy<sup>+</sup>], [spo<sup>+</sup>] et [pig<sup>+</sup>].



**Figure R-96 : Phénotype colonial de la souche *S. ambofaciens* DSM40697 sauvage et de la souche double mutante DMRL2.**

Des préparations de spores de deux souches de *S. ambofaciens* DSM40697 ont été étalées sur différents milieux de culture solides : Hickey-Tresner (HT), Milieu Minimum (MM), Milieu Minimum Mannitol (MMM), Milieu Minimum Glucose (MMG), Nutrient Agar (NA) (voir Matériel et Méthodes). Après 20 jours d'incubation à 30°C, les colonies bactériennes ont été observées à la loupe binoculaire.

Ainsi, les mutants *has*, simples et doubles, présentent un phénotype similaire à la souche sauvage quel que soit le milieu de croissance testé.

**B-V.1.2 – ACTIVITE ANTIBACTERIENNE**

Le mutant *sigB* de *S. coelicolor* A3(2) semble être affecté dans la production des antibiotiques pigmentés actinorhodine et undécylprodigiosine sur milieu MM mannitol 1M en saccharose (G. Kelemen, communication personnelle). Chez *S. ambofaciens*, la production de spiramycine, antibiotique non pigmenté, n'est pas directement décelable. Des expériences d'inhibition de croissance de la souche indicatrice *B. subtilis* ATCC6633 ont donc été réalisées afin de savoir si, sur milieu MM mannitol 1M en saccharose, les mutants *has* étaient affectés dans la production d'une activité antibactérienne.

Cependant, les zones d'inhibition observées autour des colonies des doubles mutants et du sauvage ne sont pas sensiblement différentes (Figure R-97).

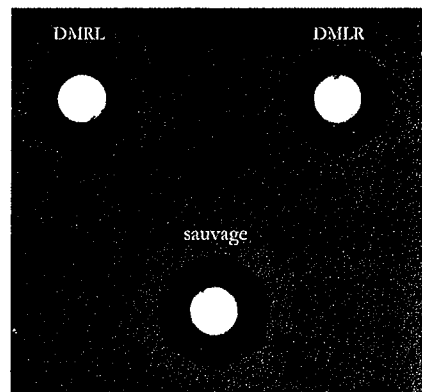


Figure R-97 : Test d'activité antibactérienne des souches sauvage et doubles mutantes *has* de *S. ambofaciens*.

Les ronds blancs correspondent à des cylindres d'agar prélevés dans une boîte de Petri où a été cultivée (jusqu'à 8 jours) à confluence une sporée de chaque souche. Le fond gris correspond à la croissance de la souche indicatrice *B. subtilis* ATCC6633. L'inhibition de croissance est révélée par la zone sombre entourant les cylindres d'agar.

## B-V.2 $\propto$ LES FACTEURS $\sigma^{Has}$ ET LA REPONSE AUX STRESS

### B-V.2.1 – RECHERCHE DE PHENOTYPES ASSOCIES AUX STRESS

La recherche d'un phénotype associé au remplacement des loci *has* a consisté en une observation qualitative du phénotype colonial dans des conditions de croissance spécifiques (stress thermique, acide-alcalin et osmotique). Des observations quantitatives ont été réalisées pour les stress ultraviolet et oxydant.

#### B-V.2.1.1 – INFLUENCE DE LA TEMPERATURE SUR LE PHENOTYPE COLONIAL

Les deux doubles mutants réciproques et la souche sauvage ont été cultivés, à 30°C et 37°C, en tapis confluent ( $10^6$  spores) sur les milieux solides HT, SFM, MM glucose et MM mannitol (la température optimale de croissance de *Streptomyces* étant de 30°C). Les phénotypes ainsi que les croissances des deux doubles mutants réciproques sont similaires à ceux de la souche sauvage quelles que soient les conditions de culture (Figure R-98).

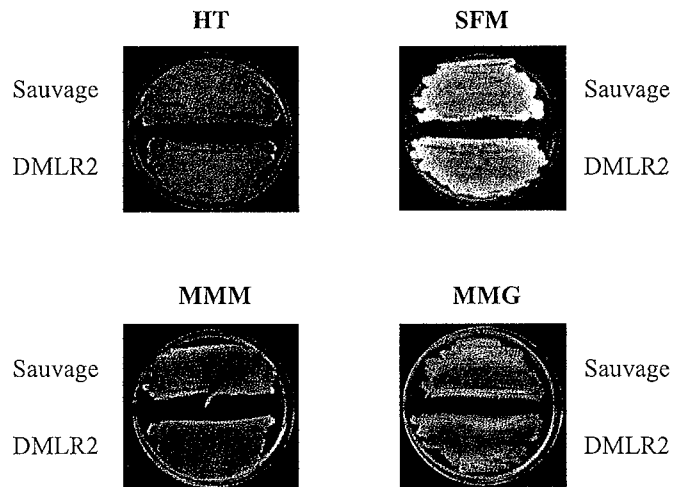


Figure R-98 : Croissance des souches de *S. ambofaciens* sauvage et DMLR2 en condition de stress thermique.

Les bactéries sont cultivées 7 jours à 37 °C, condition de stress thermique, sur les milieux Hickey-Tresner (HT), « Soya Flour Mannitol » (SFM), Milieu Minimum Glucose (MMG) et Milieu Minimum Mannitol (MMM) (voir Matériel et Méthodes).

### B-V.2.1.2 – STRESS ACIDE-ALCALIN

La croissance des *Streptomyces* est optimale lorsque le pH est proche de la neutralité. Sur milieu solide HT acide (pH 5,0) ou alcalin (pH 9,0), environ  $10^6$  spores des deux doubles mutants et de la souche sauvage ont été étalées. Aucune différence phénotypique n'a été décelée entre les souches mutantes et sauvage (Figure R-99).

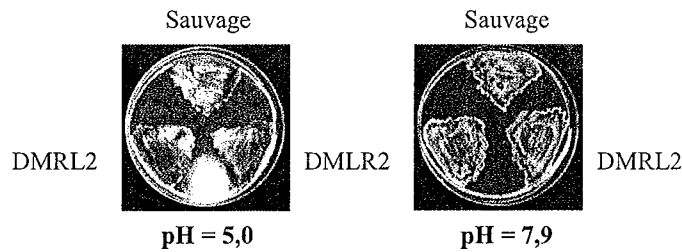


Figure R-99 : Croissance des souches de *S. ambofaciens* sauvage, DMLR2 et DMRL2 en condition de stress acide-alcalin.

Les bactéries sont cultivées 12 jours en condition de stress acide (pH 5) ou basique (pH 9) sur milieu Hickey-Tresner (HT).

### B-V.2.1.3 – STRESS OSMOTIQUE

Le facteur  $\sigma^{\text{B}}$  de *S. coelicolor* A3(2) est impliqué dans la réponse à un stress osmotique salin (Cho *et al.*, 2001). La croissance du mutant  $\sigma^{\text{B}}$  est inhibée sur milieu solide et liquide contenant 0,5 M de chlorure de potassium (KCl). Afin de tester leur comportement face au stress osmotique salin, les doubles mutants réciproques ainsi que la souche sauvage ont été cultivés (stries ou étalement de 7 à 15 colonies isolées par boîte) sur milieu NA contenant 0 M, 0,25 M, 0,5 M ou 1 M de KCl. La croissance et la différenciation des doubles mutants et du sauvage répondent de manière comparable proportionnellement à l'augmentation de la teneur en KCl (Figure R-100).

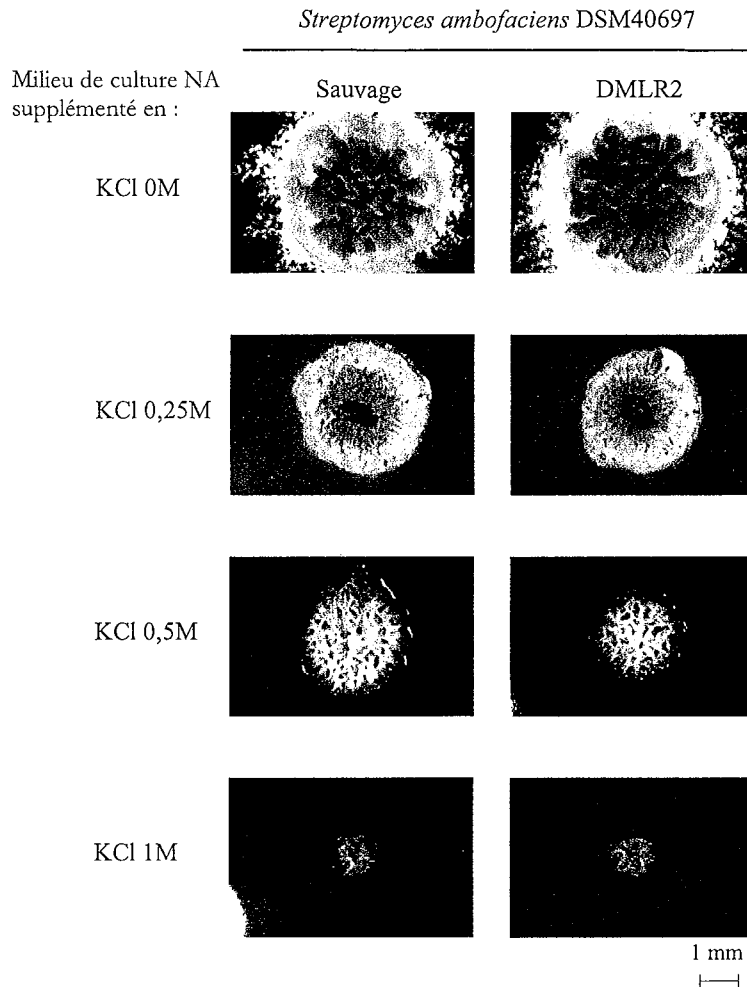


Figure R-100 : Phénotype colonial de la souche *S. ambofaciens* DSM40697 sauvage et de la souche double mutante DMLR2.

Des préparations de spores de deux souches de *S. ambofaciens* DSM40697 ont été étalées sur milieu de culture Nutrient Agar (NA), supplémenté ou non avec des doses croissantes de KCl (voir Matériel et Méthodes). Après 15 jours d'incubation à 30°C, les colonies bactériennes ont été observées à la loupe binoculaire.

#### B-V.2.1.4 – STRESS UV

La recherche du motif consensus des promoteurs potentiellement reconnus par les facteurs sigma HasL et HasR de *S. ambofaciens* dans la séquence du génome de *S. coelicolor* A3(2) a permis de mettre en évidence plusieurs gènes cibles potentiels et notamment un gène codant une protéine homologue des protéines Dps d'*E. coli*. Ces protéines sont synthétisées en réponse à des stress impliqués dans les dommages à l'ADN (Altuvia *et al.*, 1994; Martinez and Kolter, 1997). En se fixant de façon aspécifique à l'ADN, les protéines Dps formeraient une structure cristalline qui protégerait l'ADN des agressions physico-chimiques (Wolf *et al.*, 1999).

Le taux de survie des doubles mutants *bas* suite à une irradiation à différentes doses d'UV a été comparé à celui de la souche sauvage. Dans le cas d'une implication éventuelle des gènes *bas* dans le contrôle de la transcription de gènes de la famille *dps*, les doubles mutants devraient présenter une chute du taux de survie par rapport au sauvage. Les courbes de survie sont présentées Figure R-101. Aucune différence significative du taux de survie n'est observée.

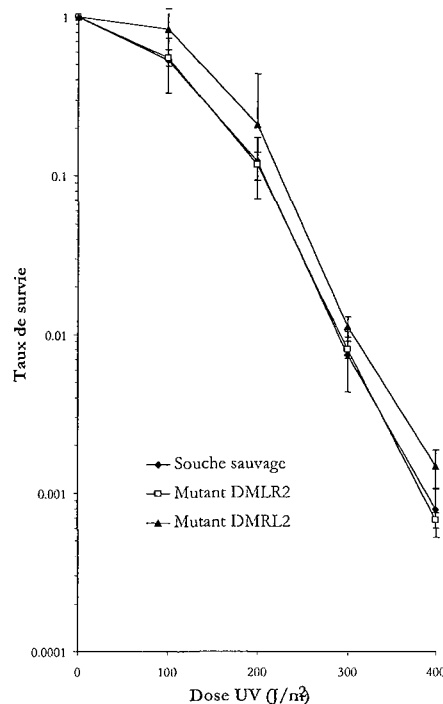


Figure R-101 : Taux de survie après irradiation UV de la souche sauvage *S. ambofaciens* DSM40697 et des doubles mutants DMLR et DMRL.

Les spores sont irradiées à différentes doses UV (en J/m<sup>2</sup>) puis incubées à l'obscurité (voir Matériel et Méthodes).

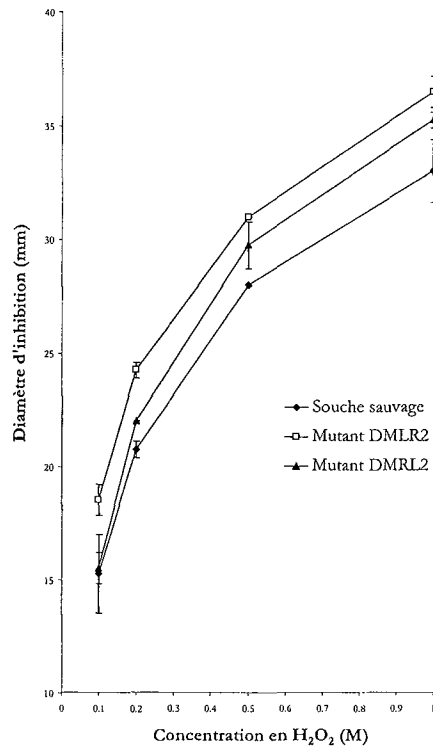
Les barres d'erreur correspondent à l'intervalle de confiance calculé avec une probabilité de 95%. Ces barres d'erreur ne sont parfois pas visibles à cause de l'échelle logarithmique.

### B-V.2.1.5 – STRESS OXYDANT

Chez *S. coelicolor* A3(2), un régulon contrôlé par le facteur  $\sigma^{\text{R}}$  appartenant à la famille des ECF est induit lors d'un stress oxydant (Paget *et al.*, 1998; Paget *et al.*, 2001b). La ménadione, diamide et peroxyde d'hydrogène (H<sub>2</sub>O<sub>2</sub>) sont de puissants agents oxydants qui ont été testés sur les doubles mutants. Aucune sensibilité particulière suite au traitement à la ménadione ou à la diamide n'a pu être décelée.



En revanche, le traitement à l' $H_2O_2$  a permis de mettre en évidence une sensibilité accrue des doubles mutants par rapport à la souche sauvage (Figure R-102). Bien que cette sensibilité soit discrète, elle apparaît significative, les expériences ayant été réalisées en triplicats.



**Figure R-102 : Inhibition de la croissance de la souche sauvage *S. ambofaciens* DSM40697 et des doubles mutants DMLR2 et DMRL2 par le peroxyde d'hydrogène.**

Les conditions de stress sont indiquées dans le Matériel et Méthodes.

Les barres d'erreur correspondent à l'intervalle de confiance calculé avec une probabilité de 95%. Ces barres d'erreur ne sont parfois pas visibles à cause de l'échelle logarithmique.

Sur l'ensemble des expériences réalisées, le diamètre d'inhibition de croissance des doubles mutants est plus grand de 11,2% ( $\pm 5,6\%$ ) que celui de la souche sauvage (qui présente une variabilité de 4,8%). Par ailleurs, les souches complémentées semblent retrouver un niveau de résistance de type sauvage (résultats non montrés).

## **B-V.2.2 – ANALYSE TRANSCRIPTIONNELLE DES GENES *HAS* EN CONDITIONS DE STRESS**

### B-V.2.2.1 – LA TRANSCRIPTION DES GENES *HASR* ET *HASL* EST INDUITE DANS DES CONDITIONS DE STRESS

Compte tenu des homologies de séquences détectées entre les facteurs  $\sigma^{Has}$  et le facteur de réponse au stress général  $\sigma^B$  de *B. subtilis*, la transcription des gènes *has* a été analysée suite à divers stress.

Ainsi, des ARN extraits à partir de cultures soumises à un stress lorsqu'elles se trouvaient en milieu de phase exponentielle de croissance ont été extraits.

Puis des expériences de cartographie à la nucléase S1 ainsi que des expériences de RT-PCR quantitative ont été réalisées à partir de ces ARN (Tableau R-11).

Ainsi, il est apparu que les stress oxydatifs (10mM H<sub>2</sub>O<sub>2</sub> pendant 30 minutes) et osmotique (500mM NaCl ou 34% saccharose pendant 30 minutes) avaient une influence directe sur la transcription des gènes *has* (Figure R-103).

En effet, en ce qui concerne le gène *hasR*, le transcrit initié au site *hasRp1* présente une expression induite d'un facteur 17 suite à un stress oxydatif, d'un facteur 25 suite à un stress salin et d'un facteur 22,5 suite à un stress osmotique par le saccharose (Figure R-103 A.).

De la même façon bien que dans une moindre mesure, le transcrit initié au site *hasLp1* pour le gène *hasL* est également induit par ces conditions, d'un facteur 7,5 pour le stress oxydatif, et 21 ou 23 pour le traitement osmotique (Figure R-103 B.).

En revanche, l'expression des transcrits initiés aux sites *hasRp2* et *hasLp2* ne semble être affectée par ces conditions de stress puisque les facteurs d'induction observés sont compris entre 1 et 2,5 au maximum.

L'expression des transcrits *has* ne semble pas être affectée par des traitements à l'éthanol (4% pendant 30 minutes) ni par un choc thermique (42°C pendant 20 minutes), à l'exception d'une induction d'un facteur 5 de l'expression du transcrit initié au site *hasLp2* suite à un choc thermique (mais cette induction serait à confirmer, compte tenu du niveau d'expression très faible de ce transcrit et donc de la fluctuation des résultats obtenus).

### B-V.2.2.2 – IMPLICATION D'AUTRES FACTEURS $\sigma$ DANS L'INDUCTION DE LA TRANSCRIPTION DES GENES *HAS* DANS DES CONDITIONS DE STRESS

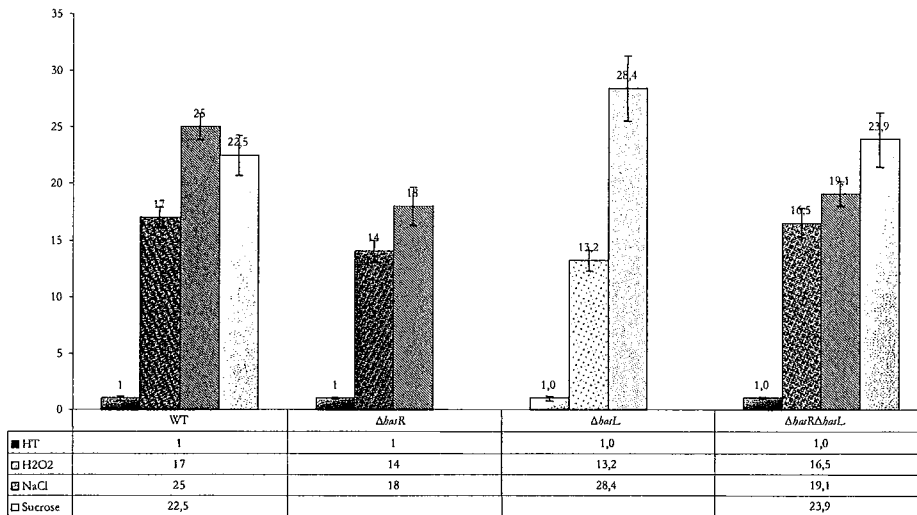
L'effet des stress oxydatif et osmotique sur la transcription des gènes *has* a également été analysé par PCR quantitative chez les simples et doubles mutants *has* (Tableau R-11). Les facteurs d'induction de l'expression des transcrits initiés aux sites *hasRp1* et *hasLp1* observés chez les mutants ne semblent être significativement différents de ceux observés chez la souche sauvage, relativement à leur niveau de transcription altéré chez ces mêmes mutants (Figure R-103).

En effet, le transcrit initié au site *hasRp1* présente une expression induite d'un facteur compris entre 13,2 et 16,5 selon la souche mutante considérée suite à un stress oxydatif, comparativement au facteur de 17 observé chez la souche sauvage. De la même manière, suite à un stress osmotique, ce transcrit est induit entre 18 et 28,4 fois chez les mutants alors qu'il l'était 22,5 ou 25 fois chez la souche sauvage (Figure R-103 A.).

En ce qui concerne le gène *hasL*, des résultats similaires ont été obtenus puisque le transcrit initié au site *hasLp1* présente une expression induite d'un facteur compris entre 6,9 et 9 selon la souche mutante considérée suite à un stress oxydatif, comparativement au facteur de 7,5 observé chez la souche sauvage. De la même manière, suite à un stress osmotique, ce transcrit est induit d'un facteur compris entre 18 et 19,5 chez les mutants alors qu'il l'était d'un facteur 21 ou 23 chez la souche sauvage (Figure R-103 B.).

Ainsi, l'induction de la transcription des gènes *has* dans des conditions de stress oxydatif et osmotique ne semble pas être dépendante des facteurs  $\sigma^{\text{Has}}$ , ou en tout cas pas exclusivement. Ces résultats suggèrent donc que des facteurs  $\sigma$  redondants pourraient contrôler l'expression des gènes *has*, notamment dans ces conditions de réponse aux stress.

A. *hasRp1*



B. *hasLp1*

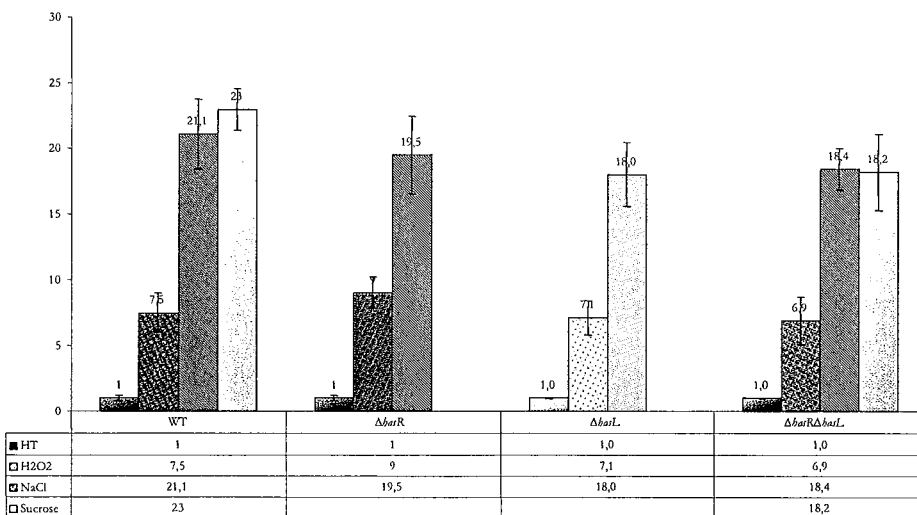


Figure R-103 : Induction de la transcription des gènes *has* suite à des stress oxydatif et salin. La quantité de chaque transcrite en milieu de phase exponentielle de croissance dans un milieu riche (HT) et sans aucun traitement a arbitrairement été fixée à 1 afin de visualiser les taux d'induction suite aux stress oxydatif et salin.

- CHAPITRE B-VI -

REGULATION DE L'EXPRESSION DES  
GENES *HASR* ET *HASL* PAR LES  
FACTEURS  $\sigma^{\text{HASR}}$  ET  $\sigma^{\text{HASL}}$  *IN VIVO*

## B-VI.1 $\oslash$ PRINCIPE DE L'ANALYSE

Afin de déterminer l'impact de l'interruption des gènes *has* sur leur propre transcription, des expériences de RT-PCR quantitative ont été réalisées. Dans cet objectif, des couples d'amorces spécifiques de chaque transcrite ont été dessinés afin d'amplifier la région promotrice conservée en amont de la cassette de résistance à l'apramycine ou spectinomycine chez les mutants (Figure R-104). En effet, la région remplacée par les cassettes Apra ou Spec chez les mutants correspond uniquement à la phase codante des gènes *has* puisque les oligonucléotides substrats de la recombinaison homologue ont pour borne 5' le codon d'initiation ou d'arrêt de la traduction des ORF *has*.

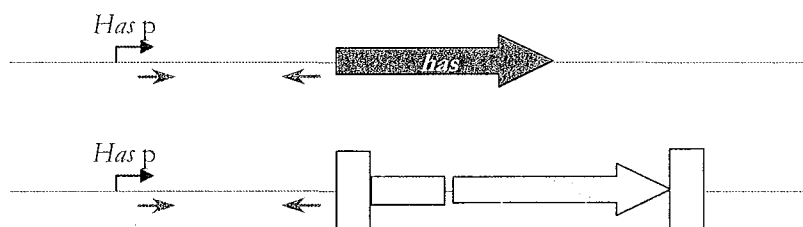


Figure R-104 : Schématisation de la localisation des amorces utilisées dans la détection, par PCR quantitative, de la transcription des gènes *has* chez les mutants.

Les couples utilisés dans ces expériences de PCR quantitative sont identiques à ceux employés précédemment pour l'analyse de la régulation de la transcription des gènes *has* chez la souche sauvage (Tableau MM-6).

Compte tenu des contraintes liées à la construction même des mutants *has*, il était nécessaire de vérifier qu'aucun transcrite initié à l'intérieur de la cassette de résistance (transcrit du gène de résistance lui-même) ne venait interférer avec la transcription résiduelle des gènes *has*. Cette hypothèse a été testée dans le cas du gène *hasL* chez un mutant  $\Delta hasL$  en comparant les données de PCR quantitative obtenues à l'aide de deux couples d'amorces (Figure R-105) :

- Le premier, L<sub>qu1</sub>/L<sub>qu2</sub>, permet de quantifier la transcription résiduelle de *hasL*, initiée aux deux promoteurs *hasLp1* et *hasLp2*.
- Le second, L<sub>qu4</sub>/L<sub>qu2</sub>, permettrait de rendre d'une transcription provenant de l'intérieur de la cassette puisque l'amorce L<sub>qu4</sub> est localisée en amont de *hasLp2*.

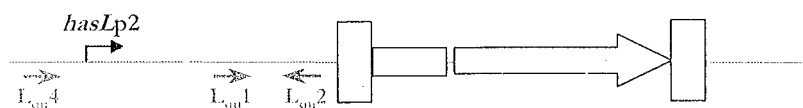


Figure R-105 : Localisation des amorces L<sub>qu1</sub>, L<sub>qu2</sub> et L<sub>qu4</sub> permettant de vérifier l'absence de transcription initiée dans la cassette de remplacement du gène *hasL*.

Aucune amplification n'a été détectée à l'aide du couple d'amorce Lqu4/Lqu2, démontrant l'absence de transcription initiée à l'intérieur de la cassette de résistance (résultats non montrés).

## B-VI.2 $\propto$ ANALYSE DE L'IMPACT DE L'INTERRUPTION DES GENES *HAS* SUR L'EXPRESSION DE CHAQUE TRANSCRIT *HAS*

Le niveau d'expression de chacun des 4 transcrits *has* a ainsi été quantifié dans les trois phases de croissance en milieu HT liquide, et ce chez trois simples mutants  $\Delta hasR$ , trois simples mutants  $\Delta hasL$  et trois double mutants  $\Delta hasR\Delta hasL$ . Les résultats bruts déduits de chaque expérience de RT-PCR quantitative sont présentés dans le Tableau R-11.

		DSM	$\Delta hasR$		$\Delta hasL$		$\Delta hasR\Delta hasL$		$\Delta hasR$ +pSET $\Omega$	$\Delta hasR$ +pSET $\Omega R$
Expo	<i>hasL</i> p2	100 / 1	113	[x1.1]	103	[x1]	264	[x2.6]	ND	ND
	<i>hasL</i> p1	100 / 9	241	[x2.4]	2591	[x26]	5761	[x58]	ND	ND
	<i>hasR</i> p2	100 / 2	368	[x3.7]	110	[x1.1]	238	[x2.4]	ND	ND
	<i>hasR</i> p1	100 / 88	292	[x2.9]	69	[/1.4]	95	[/1.05]	ND	ND
Transition	<i>hasL</i> p2	90 (0.9)	85 (0.75)	[x1]	956 (9.3)	[x10.6]	717 (2.7)	[x8]	ND	ND
	<i>hasL</i> p1	390 (3.9)	198 (0.8)	[/2]	1398 (0.5)	[x3.6]	958 (0.2)	[x2.5]	ND	ND
	<i>hasR</i> p2	119 (1.2)	1263 (3.4)	[x10.6]	105 (0.95)	[/1.1]	1434 (6)	[x12]	ND	ND
	<i>hasR</i> p1	3000 (30)	1559 (5.3)	[/1.9]	2432 (35.2)	[/1.2]	876 (9.2)	[/3.4]	ND	ND
Stat	<i>hasL</i> p2	128 (1.3)	28 (0.25)	[/4.6]	1277 (12.4)	[x10]	363 (1.4)	[x2.8]	35 [/3.7]	103 [/1.2]
	<i>hasL</i> p1	1100 (11)	58 (0.25)	[/19]	607 (0.25)	[/1.8]	532 (0.1)	[/2]	53 [/20.7]	816 [/1.3]
	<i>hasR</i> p2	229 (2.3)	609 (1.65)	[x2.7]	389 (3.5)	[x1.7]	247 (1)	[x1.1]	586 [x2.6]	182 [/1.3]
	<i>hasR</i> p1	10900 (109)	368 (1.3)	[/29.6]	4230 (61.3)	[/2.6]	320 (3.4)	[/34]	331 [/32.9]	9483 [/1.15]
H <sub>2</sub> O <sub>2</sub>	<i>hasL</i> p2	96 (1)	136 (1.2)	[x1.4]	100 (1)	[x1]	291 (1.1)	[x3]	ND	ND
	<i>hasL</i> p1	745 (7.5)	2172 (9)	[x2.9]	18402 (7.1)	[x24.7]	39746 (6.9)	[x53.3]	ND	ND
	<i>hasR</i> p2	108 (1.1)	740 (2)	[x6.8]	107 (1)	[x1]	382 (1.6)	[x3.5]	ND	ND
	<i>hasR</i> p1	1698 (1.7)	4090 (14)	[x2.4]	910 (13.2)	[/1.9]	1569 (16.5)	[/1.1]	ND	ND
NaCl	<i>hasL</i> p2	168 (1.7)	110 (1)	[/1.5]	93 (0.9)	[/1.8]	320 (1.2)	[x1.9]	ND	ND
	<i>hasL</i> p1	2108 (21)	4702 (19.5)	[x2.2]	46600 (18)	[x22.1]	106000 (18.4)	[x50.3]	ND	ND
	<i>hasR</i> p2	248 (2.5)	664 (1.8)	[x2.7]	354 (3.2)	[x1.4]	417 (1.75)	[x1.7]	ND	ND
	<i>hasR</i> p1	2497 (25)	5261 (18)	[x2.1]	1963 (28.4)	[/1.3]	1811 (19)	[/1.4]	ND	ND
Sucrose	<i>hasL</i> p2	ND	ND	ND	ND	ND	ND	ND	ND	ND
	<i>hasL</i> p1	2297 (23)	ND	ND	ND	ND	104850 (18.2)	[x49.7]	ND	ND
	<i>hasR</i> p2	ND	ND	ND	ND	ND	ND	ND	ND	ND
	<i>hasR</i> p1	2253 (22.5)	ND	ND	ND	ND	2274 (24)	[/1.1]	ND	ND

Tableau R-11 : Niveaux d'expression des transcrits *has* dans diverses conditions et chez diverses souches déduits des analyses en RT-PCR quantitative.

Deux valeurs sont indiquées pour la souche sauvage en phase exponentielle de croissance : la première est fixée arbitrairement à 100 et sera celle utilisée comme référence pour toutes les autres valeurs du tableau ; la seconde correspond à l'abondance relative de chaque transcrit si l'expression globale des deux gènes *has* est de 100.

Les valeurs entre parenthèses correspondent aux facteurs d'induction observés au sein de chaque souche. Pour la quantification au cours de la croissance, ce facteur correspond à l'induction observée par rapport à la phase exponentielle. Pour les conditions de stress, ce facteur correspond à l'induction du gène par rapport à une croissance normale en phase exponentielle.

Pour les différents mutants la valeur entre crochets représente la modification de l'expression (X pour induction et / pour répression) par rapport à la souche sauvage.

Chaque valeur est une valeur moyenne de 3 expérimentations indépendantes (cf Matériel et Méthodes). Pour les différents mutants, la valeur est une moyenne de celles obtenues chez plusieurs représentants indépendants : 3 pour  $\Delta hasR$ ,  $\Delta hasL$  et  $\Delta hasR\Delta hasL$ , et 2 pour  $\Delta hasR$ +pSET $\Omega$  et  $\Delta hasR$ +pSET $\Omega R$ .

Les résultats obtenus peuvent être présentés sous plusieurs formes selon les interprétations qui peuvent en être déduites.

En effet, nous pouvons tout d'abord considérer les résultats obtenus pour les quatre transcrits chez une souche mutante en particulier et ainsi en déduire le rôle du facteur  $\sigma^{\text{Has}}$  (interrompu dans la souche considérée) sur chacun des différents promoteurs.

Nous pouvons également présenter les résultats en fonction de la phase de croissance puisqu'il s'est avéré que l'effet d'un facteur  $\sigma^{\text{Has}}$  sur un promoteur donné était dépendant de la phase de croissance considérée.

Une représentation intéressante des données consiste à compiler les résultats obtenus pour un promoteur donné dans toutes les souches et dans toutes les phases de croissance analysées ce qui permet de schématiser l'impact de l'interruption de chacun ou des deux gènes *has* sur l'évolution de l'expression du transcrit considéré au cours de la croissance (Figures R-106 et R-107).

Les résultats vont ici être présentés dans le détail en fonction du type d'impact de l'interruption d'un gène *has* sur les niveaux de transcription étudiés. Ainsi, l'interruption d'un gène *has* peut avoir pour résultante la diminution de la quantité de transcrit par rapport à la souche sauvage ou au contraire une augmentation de cette quantité de transcrit. Afin de clarifier l'exposé des résultats, cette partie sera plus interprétative et exposera les types de régulation transcriptionnelle que suggèrent les données obtenues. Les niveaux plus complexes de régulation, faisant notamment intervenir les anti-facteurs  $\sigma$ , seront abordés dans la partie Discussion.

Les résultats obtenus pour les transcrits *hasRp2* et *hasLp2* ne seront pas détaillés. En effet, compte-tenu du faible niveau d'expression de ces transcrits chez la souche sauvage, les valeurs des Ct obtenues sont élevées et présentent une fluctuation plus importante, reflétant vraisemblablement les limites de la technique de PCR quantitative pour la détection de faibles quantités de transcrits. Ainsi, les variations observées chez les mutants restent également dans des faibles niveaux d'expression et donc présentent donc des écarts-types élevés (Figure R-106). De plus, les variations du niveau d'expression du transcrit *hasLp2*, observées notamment chez les mutants  $\Delta\textit{hasL}$  et  $\Delta\textit{hasR}\Delta\textit{hasL}$  en phase de transition et en phase stationnaire de croissance, ne peuvent être interprétées par le contrôle direct des facteurs  $\sigma^{\text{Has}}$  puisque aucun promoteur putatif de type  $\sigma^{\text{Has}}$  n'a été identifié pour ce transcrit. Seule une cascade de régulation faisant intervenir d'autres facteurs  $\sigma$  que les facteurs  $\sigma^{\text{Has}}$ , pourrait permettre d'expliquer ces résultats.



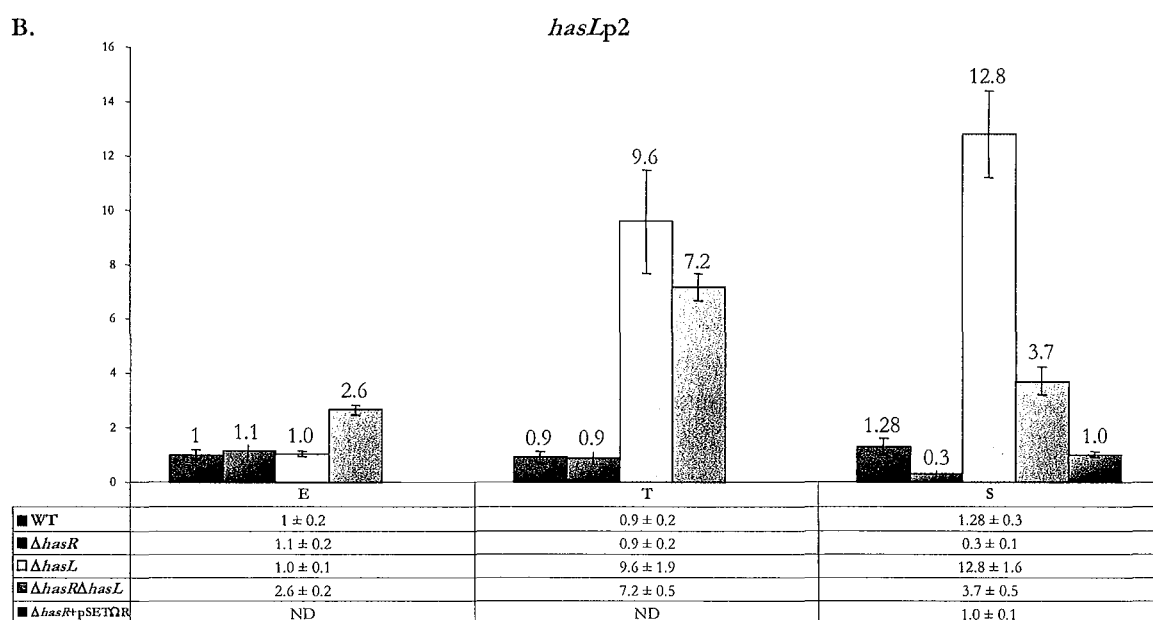
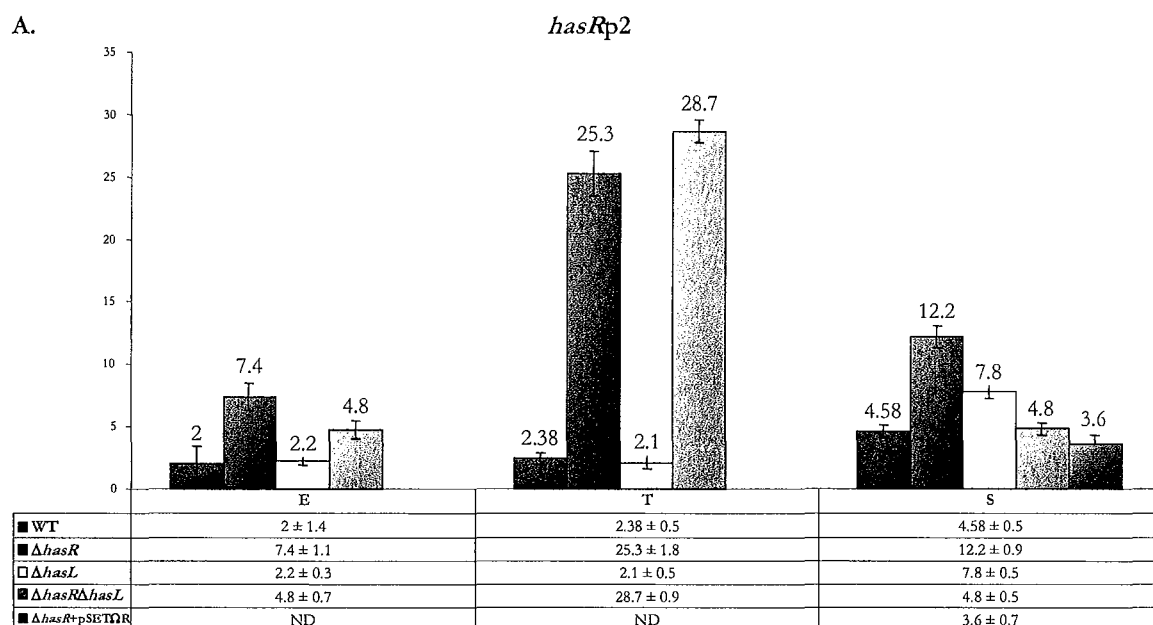


Figure R-106 : Niveaux d'expression des transcrits initiés aux sites *hasRp2* (A.) et *hasLp2* (B.) au cours de la croissance chez la souche sauvage *S. ambotaciensis* DSM40697, chez les simples et doubles mutants *has*, ainsi que chez le complémentant.

Il est à noter que les échelles des différents graphiques sont différentes. Ces graphiques sont réalisés à partir des valeurs du tableau R-9, dans lequel le niveau de transcription global des gènes *has* est arbitrairement fixé à 100.

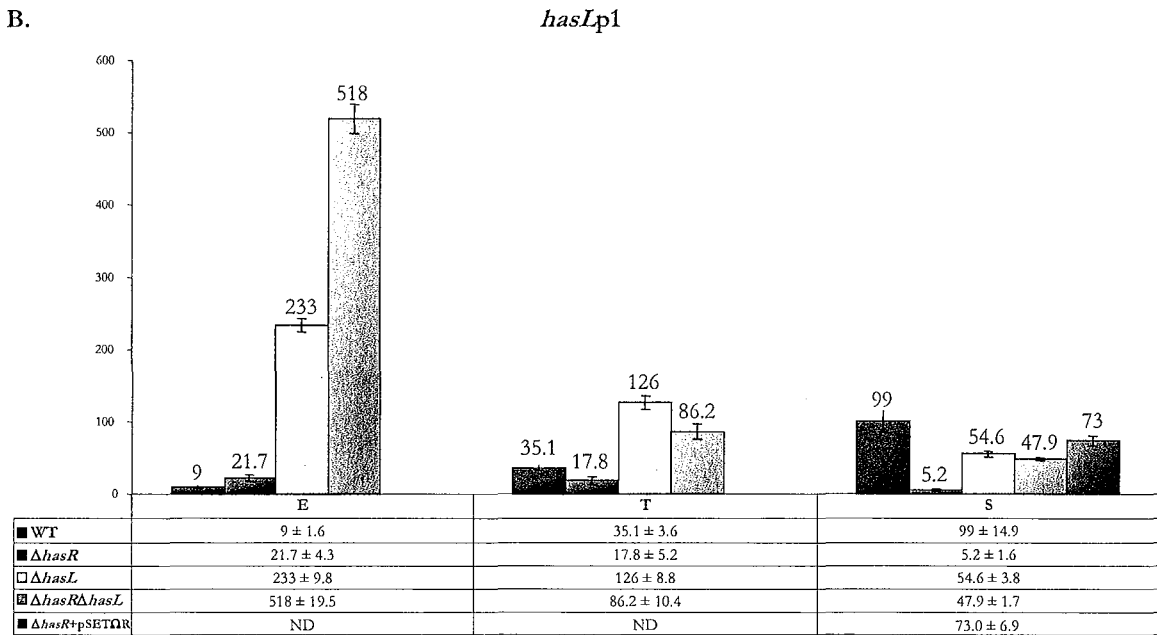
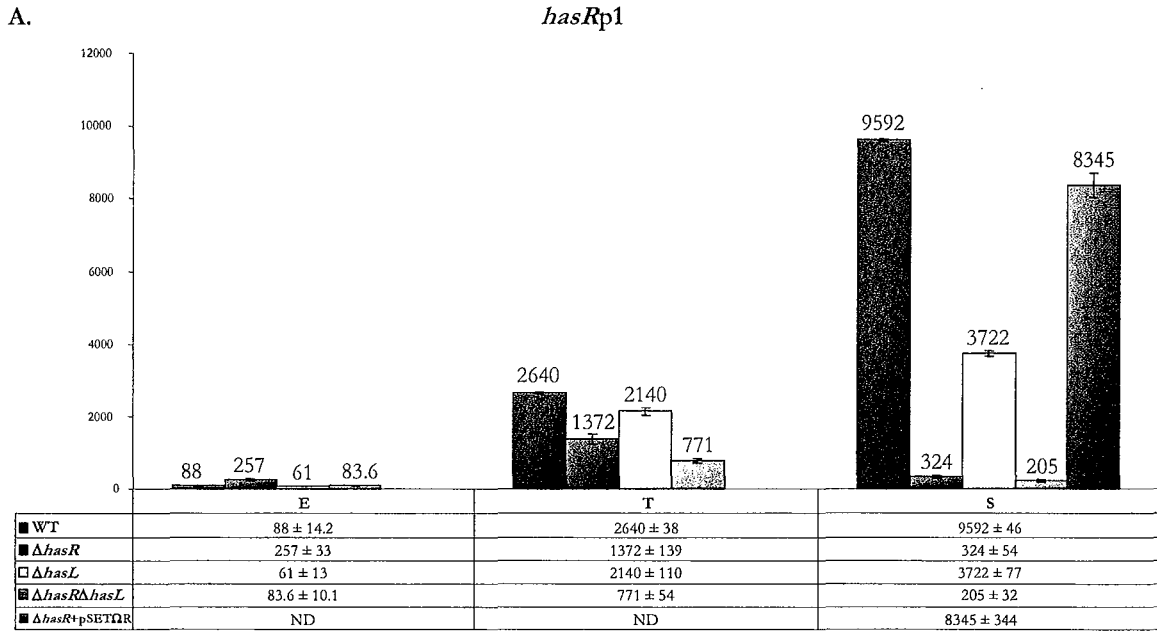


Figure R-107 : Niveaux d'expression des transcrits initiés aux sites *hasRp2* (A.) et *hasLp2* (B.) au cours de la croissance chez la souche sauvage *S. ambofaciens* DSM40697, chez les simples et doubles mutants *has*, ainsi que chez le complémentant.

Il est à noter que les échelles des différents graphiques sont différentes. Ces graphiques sont réalisés à partir des valeurs du tableau R-9, dans lequel le niveau de transcription global des gènes *has* est arbitrairement fixé à 100.

Cet effet « inhibiteur » est également suggéré dans le cas du facteur  $\sigma^{\text{HasL}}$  puisque il a été montré que le transcrite *hasLp1* est 26 fois plus exprimé dans une souche mutante  $\Delta\text{hasL}$  en phase exponentielle de croissance que dans la souche sauvage, et 3,6 fois en phase de transition.

L'ensemble de ces résultats suggèrent donc que les facteurs  $\sigma^{\text{Has}}$  s'auto-régulent. Cet auto-contrôle peut être positif, dans le sens où les facteurs  $\sigma^{\text{Has}}$  induisent leur propre transcription, mais également négatif. Cette « inhibition » de la transcription pourraient résulter d'une compétition entre les facteurs  $\sigma^{\text{Has}}$  et d'autres facteurs sigma (voir Discussion).

### B-VI.2.2 – EFFET DE L'INTERRUPTION D'UN GÈNE *HAS* SUR LA TRANSCRIPTION DE L'AUTRE COPIE *HAS*

La quantification du niveau d'expression d'un gène *has* dans la souche simple mutante interrompue pour le second gène *has* a permis de révéler des phénomènes de « cross-régulation ». Comme précédemment, cette régulation croisée peut être « activatrice ou « inhibitrice ».

En effet, la quantité de transcrite *hasRp1* est 3 fois moindre chez un mutant  $\Delta\text{hasL}$  en phase stationnaire de croissance que chez la souche sauvage, suggérant que le gène *hasR* est régulé par le facteur  $\sigma^{\text{HasL}}$  en phase stationnaire de croissance chez la souche sauvage (Figure R-110).

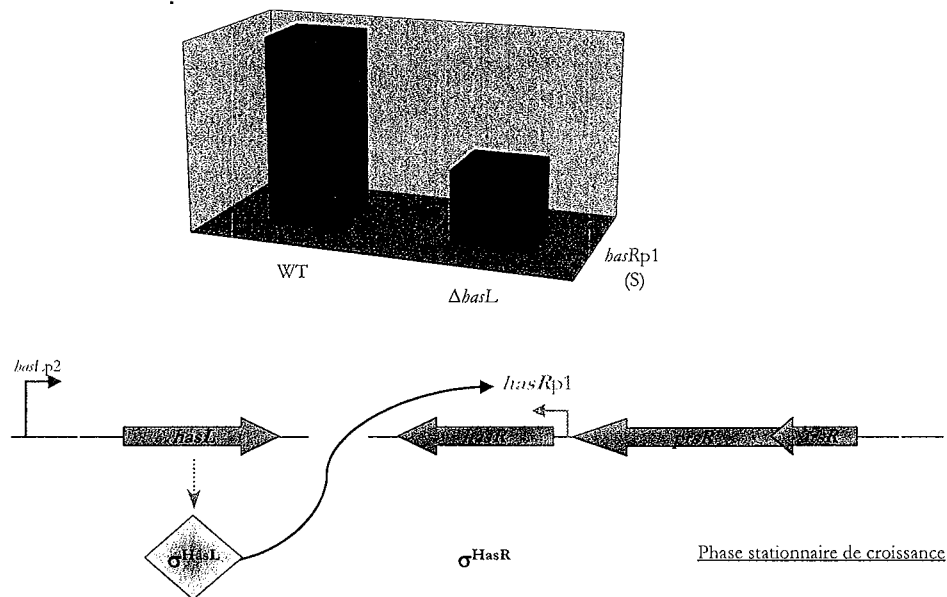


Figure R-110 : Interruption d'un gène *has* et diminution du niveau de transcription de l'autre copie *has*: exemple du transcrite *hasRp1* chez un mutant  $\Delta\text{hasL}$  en phase stationnaire de croissance.

En haut : Niveaux d'expression du transcrite *hasRp1* chez la souche sauvage et chez un mutant  $\Delta\text{hasL}$  en phase stationnaire de croissance (quantifiés par RT-PCR quantitative).

En bas : schématisation de l'interprétation des données en termes de régulation transcriptionnelle : contrôle croisé du gène *hasR* par le facteur  $\sigma^{\text{HasL}}$ .

Il est cependant important de noter que ces analyses ne permettent pas de clairement définir l'influence relative de chacun des deux facteurs  $\sigma^{\text{Has}}$ . En effet, dans le cas du transcrit *hasRp1*, les seules données obtenues chez les simples mutants  $\Delta\textit{hasR}$  suggèreraient une régulation quasi exclusive par le facteur  $\sigma^{\text{HasR}}$  en phase stationnaire de croissance, tandis que les résultats obtenus chez les souches  $\Delta\textit{hasL}$  tendent à montrer une intervention équivalente de chacun des deux facteurs  $\sigma^{\text{Has}}$ . La combinaison de ces deux types de données permettent alors d'émettre l'hypothèse selon laquelle le gène *hasL* serait lui aussi régulé par le facteur  $\sigma^{\text{HasR}}$  en phase stationnaire de croissance et qu'ainsi, non seulement la transcription par le facteur  $\sigma^{\text{HasR}}$  serait inhibée dans un simple mutant  $\Delta\textit{hasR}$  mais également une partie de la régulation via le facteur  $\sigma^{\text{HasL}}$ , d'où le très faible niveau de transcription résiduelle du transcrit *hasRp1* chez ce type de mutant. Des hypothèses alternatives faisant intervenir des niveaux de régulation plus complexes, avec notamment les anti-facteur  $\sigma$ , pourraient également permettre d'expliquer ces résultats (voir Discussion).

Le niveau de transcription du gène *hasL* diminue d'un facteur 19 chez les souches simple mutantes  $\Delta\textit{hasR}$  en phase stationnaire de croissance comparativement à la souche sauvage. Cette diminution du niveau de transcription du gène *hasL* dans un mutant  $\Delta\textit{hasR}$  suggère la « cross-régulation » du gène *hasL* par le facteur  $\sigma^{\text{HasR}}$  (Figure R-111) mais peut-être également la capacité amoindrie du facteur  $\sigma^{\text{HasL}}$  à s'auto-réguler, de part sa plus faible expression.

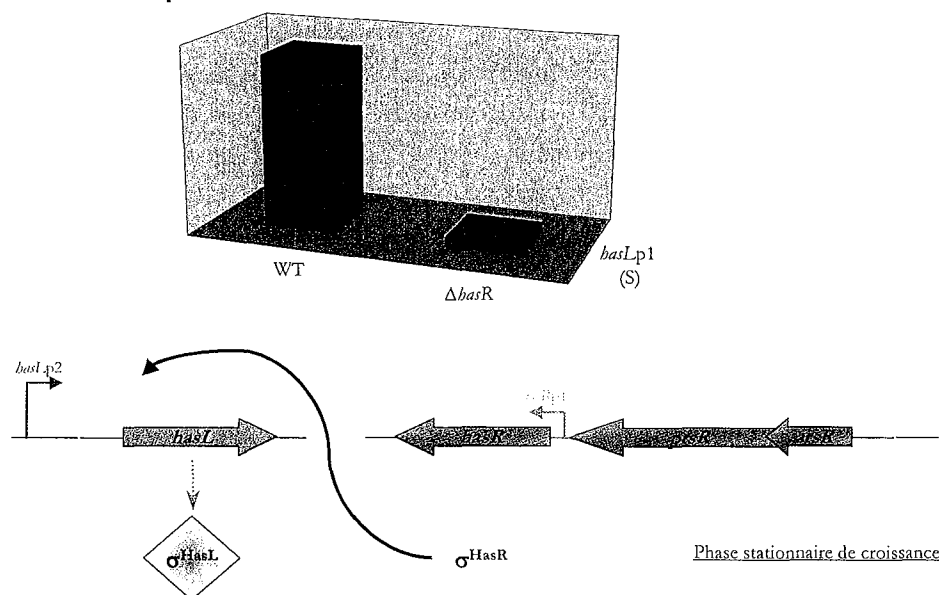


Figure R-111 : Interruption d'un gène *has* et diminution du niveau de transcription de l'autre copie *has*: exemple du transcrit *hasRp1* chez un mutant  $\Delta\textit{hasR}$  en phase stationnaire de croissance.

En haut : Niveaux d'expression du transcrit *hasRp1* chez la souche sauvage et chez un mutant  $\Delta\textit{hasR}$  en phase stationnaire de croissance (quantifiés par RT-PCR quantitative).

En bas : schématisation de l'interprétation des données en termes de régulation transcriptionnelle : contrôle croisé du gène *hasL* par le facteur  $\sigma^{\text{HasR}}$ .

Ces phénomènes de « cross-régulation » ont également pu être déduits des résultats obtenus :

- pour le transcrit *hasRp1* en phase exponentielle de croissance (transcription 1,4 fois moindre dans un mutant  $\Delta hasL$ ) et en phase de transition (transcription 1,2 fois moindre dans un mutant  $\Delta hasL$ ) ;
- et pour le transcrit *hasLp1* en phase de transition de croissance (transcription 2 fois moindre dans un mutant  $\Delta hasR$ ).

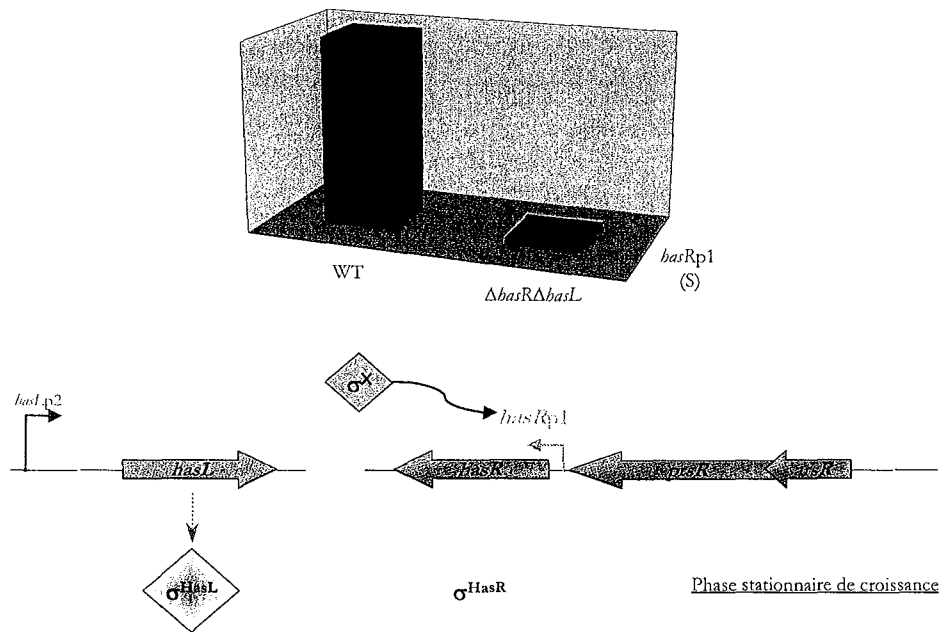
Enfin, un effet « inhibiteur » croisé a été montré dans le cas du transcrit *hasLp1* qui est 2,4 fois plus exprimé dans une souche mutante  $\Delta hasR$  en phase exponentielle de croissance que dans la souche sauvage.

L'ensemble de ces résultats suggère donc que les facteurs  $\sigma^{Has}$  se régulent de façon croisée. A l'image de l'auto-régulation, cette « cross-régulation » peut être positive ou négative.

### **B-VI.2.3 – EFFET DE L'INTERRUPTION DES DEUX GÈNES *HAS***

Il est important de noter l'effet additionnel de l'interruption des deux gènes *has* chez les doubles mutants  $\Delta hasR\Delta hasL$ . En effet, lorsqu'une diminution du niveau d'expression d'un transcrit est observée dans chacun des deux simple mutants  $\Delta hasR$  et  $\Delta hasL$ , cette diminution est d'autant plus élevée chez les doubles mutants  $\Delta hasR\Delta hasL$  (cas du transcrit *hasRp1* en phase de transition et stationnaire et du transcrit *hasLp1* en phase stationnaire, Figure R-107). Par ailleurs, lorsque l'interruption de chaque gène *has* a un effet opposé sur le niveau d'expression d'un même transcrit (cas des transcrits *hasRp1* en phase exponentielle et *hasLp1* en phase de transition, Figure R-103), le niveau de transcription observé chez un double mutant  $\Delta hasR\Delta hasL$  est intermédiaire. Enfin, lorsque l'interruption de chaque gène *has* était accompagnée de l'augmentation du niveau d'un transcrit (cas du transcrit *hasLp1* en phase exponentielle, Figure R-107), cette augmentation est d'autant plus élevée chez un double mutant  $\Delta hasR\Delta hasL$ .

Par ailleurs la détection d'une expression résiduelle des transcrits *has* chez une souche où les deux gènes *has* ont été interrompus suggère l'intervention d'autres facteurs sigma dans la régulation des gènes *has*. Par exemple, dans une souche double mutante  $\Delta hasR\Delta hasL$  en phase stationnaire de croissance, une transcription résiduelle du transcrit *hasRp1* a pu être détectée (Figure R-112).



**Figure R-112 : Interruption des deux gènes *has* et détection d'une expression résiduelle d'un transcrit *has* : exemple du transcrit *hasRp1* chez un mutant  $\Delta hasR\Delta hasL$  en phase stationnaire de croissance.**

En haut : Niveaux d'expression du transcrit *hasRp1* chez la souche sauvage et chez un mutant  $\Delta hasR\Delta hasL$  en phase stationnaire de croissance (quantifiés par RT-PCR quantitative).

En bas : schématisation de l'interprétation des données en termes de régulation transcriptionnelle : contrôle du gène *hasR* par d'autres facteurs  $\sigma$  ( $\sigma^X$ ).

Ces résultats sont compatibles avec le maintien de l'induction des gènes *has* en conditions de stress chez les doubles mutants (voir paragraphe B-V.2.2.2) et soulignent à nouveau la redondance des facteurs  $\sigma$ .

#### **B-VI.2.4 – COMPLEMENTATION FONCTIONNELLE**

Le niveau d'expression des quatre transcrits *has* a été quantifié chez deux complémentants du gène *hasR* en phase stationnaire de croissance (c'est à dire deux transconjugants issus du simple mutant  $\Delta hasR$  R2 ayant intégré un vecteur recombinant pSET $\Omega$ R renfermant une copie sauvage du gène *hasR*). Cette quantification a également été réalisée chez un transconjugant issu du même simple mutant  $\Delta hasR$  R2 ayant intégré un vecteur recombinant pSET $\Omega$  vide, afin de vérifier que cette construction n'influe pas la transcription des gènes *has*.

Comme pour l'analyse de la transcription chez les mutants, seules les variations observées pour les transcrits *hasRp1* et *hasLp1* seront détaillées.

Ainsi, alors que les niveaux d'expression des transcrits chez la souche NSApSET $\Omega$  ne sont pas apparus significativement différents de ceux observés chez la souche simple mutante  $\Delta$ *basR* R2, il est apparu que chez les souches complémentantes NSApSET $\Omega$ R analysées, l'expression des transcrits *basRp1* et *basLp1* était restaurée, avec un niveau quasi sauvage (restauration à 87% et 74% respectivement).

## ❧ DISCUSSION ❧



**A - PREMIERE PARTIE :**

**COMPLEXITE DE LA REGULATION DE L'EXPRESSION GENIQUE**

**CHEZ LES *STREPTOMYCES* \_\_\_\_\_ 255**

**- CHAPITRE A-I -**

**LES DIVERS NIVEAUX DE REGULATION DE L'EXPRESSION DES GENES CODANT LES FACTEURS SIGMA CHEZ LES *STREPTOMYCES***

**A-I.1 ☞ Le contrôle transcriptionnel des gènes *has* \_\_\_\_\_ 257**

A-I.1.1 – La transcription des gènes *has* est sous contrôle de promoteurs multiples \_\_\_\_\_ 257

A-I.1.2 – Les facteurs sigma Has et la répression transcriptionnelle \_\_\_\_\_ 258

A-I.1.3 – Modulation de la régulation croisée des gènes *has* en fonction du stade de croissance \_\_\_\_\_ 258

**A-I.2 ☞ L'activité des facteurs sigma Has est-elle régulée par des mécanismes post-traductionnels ? \_\_\_\_\_ 261**

A-I.2.1 - Synthèse d'un facteur sigma sous forme de propolypeptide (pro-sigma) \_\_\_\_\_ 261

A-I.2.2 - La protéolyse, mécanisme de contrôle de la différenciation cellulaire \_\_\_\_\_ 262

A-I.2.3 - Contrôle de l'activité par des inhibiteurs (anti-facteur sigma) \_\_\_\_\_ 263

**- CHAPITRE A-II -**

**LA REPONSE AU STRESS CHEZ LES *STREPTOMYCES***

**A-II.1 ☞ La redondance des systèmes de régulation transcriptionnelle \_\_\_\_\_ 266**

**A-II.2 ☞ Les facteurs sigma : un lien entre différenciation morphologique et réponse au stress \_\_\_\_\_ 267**

## B - SECONDE PARTIE :

### CREATION DE DIVERSITE GENETIQUE PAR REARRANGEMENT

D'ADN \_\_\_\_\_ 268

#### - CHAPITRE B-I -

##### POLYMORPHISME DES EXTREMITES DES REPLICONS LINEAIRES

B-I.1	œ Distribution de la linéarité chromosomique chez les bactéries	270
B-I.2	œ Réplicons linéaires et réarrangements chromosomiques	272
B-I.2.1	– La linéarité chromosomique confère-t-elle un avantage ?	272
B-I.2.1.1	– Régiospécificité des réarrangements chez les <i>Streptomyces</i>	272
B-I.2.1.2	– La linéarité chromosomique favorise-t-elle l'instabilité chromosomique ?	273
B-I.2.2	- Mécanismes moléculaires des réarrangements chromosomiques	274
B-I.2.3	- réarrangements chromosomiques et évolution rapide des génomes	275
B-I.2.3.1	- Recombinaison subtélomérique et acquisition d'information génétique	276
B-I.2.3.1.1	- Transfert horizontal d'information génétique chez les <i>Streptomyces</i>	276
B-I.2.3.1.2	– Mécanismes de transfert d'information génétique aux extrémités des chromosomes des <i>Streptomyces</i>	279
B-I.2.3.1.3	– Interactions plasmide/chromosome et acquisition de la linéarité chromosomique	281
B-I.2.3.1.2.1	- Interaction entre un plasmide linéaire et un chromosome circulaire ancestral chez <i>S. coelicolor</i> et <i>B. burgdorferi</i>	281
B-I.2.3.1.2.2	- Acquisition de fonctions chromosomiques sur un plasmide linéaire ancestral chez <i>Agrobacterium tumefaciens</i>	282
B-I.2.3.2	– Les extrémités terminales inversées des chromosomes de <i>Streptomyces</i> ont-elles un rôle ?	283
B-I.2.3.2.1	- Implication de l'instabilité dans l'évolution rapide des extrémités des <i>Streptomyces</i>	283
B-I.2.3.2.2	– Les TIR : un substrat pour la recombinaison terminale ?	284

#### - CHAPITRE B-II -

##### DUPLICATIONS CHROMOSOMIQUES ET EVOLUTION

B-II.1	œ La dynamique des génomes	287
B-II.1.1	– Les <i>Streptomyces</i> : modèle d'étude de l'évolution des génomes	287

B-II.1.2 – Evolution des génomes : transfert horizontal et duplication génique	288
<b>B-II.2 <math>\propto</math> Le devenir des gènes dupliqués <i>hasR</i> et <i>hasL</i> au cours de l'évolution</b>	<b>291</b>
B-II.2.1 – Un mécanisme d'évolution concertée à l'origine de la conservation de la duplication des gènes <i>has</i> ?	291
B-II.2.2 – L'avantage évolutif de la duplication des gènes <i>has</i>	292

## C - TROISIEME PARTIE :

<b>PERSPECTIVES</b>	<b>295</b>
---------------------	------------

<b>C-I <math>\propto</math> Valeurs adaptatives des duplications terminales</b>	<b>296</b>
<b>C-II <math>\propto</math> Contrôle de la réponse au stress chez les <i>Streptomyces</i></b>	<b>298</b>
C-II.1 – Contrôle de l'expression des gènes <i>has</i>	298
C-II.1.1 – Contrôle la transcription des gènes <i>has</i>	298
C-II.1.2 – Les anti-facteurs $\sigma$ contrôlant l'activité des facteurs $\sigma^{\text{Has}}$	299
C-II.2 – Le rôle du régulon $\sigma^{\text{Has}}$	299
C-II.2.1 – Recherche de phénotype chez des mutants <i>has</i>	299
C-II.2.2 – Identification des gènes cibles des facteurs $\sigma^{\text{Has}}$	300

A - PREMIERE PARTIE :

COMPLEXITE DE LA REGULATION  
DE L'EXPRESSION GENIQUE CHEZ  
LES *STREPTOMYCES*

- CHAPITRE A-I -

LES DIVERS NIVEAUX DE REGULATION  
DE L'EXPRESSION DES GENES CODANT  
LES FACTEURS SIGMA CHEZ LES  
*STREPTOMYCES*

## A-I.1 $\otimes$ LE CONTROLE TRANSCRIPTIONNEL DES GENES *HAS*

### A-I.1.1 – LA TRANSCRIPTION DES GENES *HAS* EST SOUS CONTROLE DE PROMOTEURS MULTIPLES

Ce travail a permis de suggérer un modèle de régulation des gènes *has*. Deux niveaux d'expression peuvent être distingués dans ce modèle :

- un niveau d'expression basal initié aux promoteurs *hasRp2* et *hasLp2*
- un niveau d'expression inductible au cours de la croissance avec les transcrits initiés aux sites *hasRp1* et *hasLp1*. Ces promoteurs de type  $\sigma^B$  semblent être reconnus par les facteurs  $\sigma^{Has}$  eux-mêmes (avec deux niveaux de régulation : une auto régulation et une régulation croisée) mais aussi par d'autres facteurs  $\sigma$  de type  $\sigma^B$ .

Chez *E. coli*, le gène *rpoH* codant le facteur  $\sigma^{32}$  (spécifique de la réponse au choc thermique ou "heat shock") constitue un exemple de gène codant un facteur  $\sigma$  où les différents niveaux de régulation, notamment la transcription, ont été élucidés (Yura *et al.*, 1993). Ainsi, au moins 4 promoteurs distincts ont été caractérisés en amont de ce gène : ils sont sous le contrôle de facteurs  $\sigma$  de la famille  $\sigma^{70}$  dont un  $\sigma^{ECF}$  (codé par *sigE*). Ce dernier serait impliqué dans la tolérance à haute température nécessitant l'intervention des produits du régulon, notamment les protéines chaperonnes.

La présence de promoteurs multiples en amont de gènes de *Streptomyces* n'est pas rare, notamment dans le cas des gènes codant les facteurs  $\sigma$ .

En effet, 4 promoteurs ont été identifiés en amont du gène *sigH* codant un facteur  $\sigma$  impliqué dans la différenciation et dans la réponse aux stress (voir paragraphe B-II.2.3.1.2 de l'introduction, Kormanec *et al.*, 2000; Kelemen *et al.*, 2001). Ces quatre promoteurs permettent l'expression différentielle du gène *sigH* en fonction des conditions de stress ou des étapes de la différenciation. En effet, alors que le promoteur *sigHp1* est exprimé constitutivement, l'activité du promoteur *sigHp2* est induite en conditions de stress salin. De plus, la transcription initiée au promoteur *sigHp3* est activée au cours de la différenciation et induite suite à un choc thermique. Enfin, le promoteur *sigHp4* est légèrement activé par divers stress (stress salin, éthanol, choc thermique) et son activité au cours de la différenciation est l'opposée de celle de *sigHp3*, c'est à dire qu'il est particulièrement induit en phase végétative de croissance.

Les phénomènes d'auto-contrôle transcriptionnel caractérisés au cours de ce travail pour les gènes *has* ont également été décrits dans le cas du gène *sigH*. En effet, dans un mutant interrompu pour le gène *sigH*, alors que l'activité des promoteurs *sigHp1*, *sigHp3* et *sigHp4* ne semble pas altérée, la transcription initiée au promoteur *sigHp2* est très largement diminuée, indiquant donc une auto-activation de *sigH* via son promoteur *sigHp2* (Sevcikova *et al.*, 2001)

### A-I.1.2 – LES FACTEURS SIGMA HAS ET LA REPRESSION TRANSCRIPTIONNELLE

L'analyse de la transcription des gènes *bas* chez les mutants a également permis de révéler des phénomènes d'inhibition de la transcription médiés par un facteur  $\sigma$  (voir paragraphe V.2.2.3 de la seconde partie des Résultats). Une hypothèse permettant d'expliquer ce type de processus consisterait à évoquer l'existence de phénomènes de compétition entre deux (ou plus) facteurs  $\sigma$ . Cette compétition pourrait exister de part des différences d'affinité pour un même promoteur : un facteur  $\sigma^{\text{Has}}$  pourrait par exemple empêcher un autre facteur  $\sigma$  de reconnaître son promoteur et ainsi diminuer sa propre transcription, comme les résultats de PCR quantitative chez les mutants le suggèrent.

Un tel mécanisme de répression transcriptionnelle a été caractérisé dans le cas des facteurs  $\sigma^5$  et  $\sigma^{70}$  chez *E. coli*, qui entre en compétition pour une quantité limitée du core enzyme de l'ARN polymérase durant la phase stationnaire de croissance (Farewell *et al.*, 1998). Cette compétition peut être simple si seules les quantités respectives de chaque facteur  $\sigma$  entre en jeu mais elle est réellement plus complexe dans le sens où les affinités de chaque facteur  $\sigma$  pour le core enzyme ne sont pas identiques, au moins *in vitro*.

Ce phénomène d'inhibition de la transcription par un facteur  $\sigma$  a également été suggéré dans le cas des facteurs  $\sigma$  codés par les gènes *sigH*, *sigI* et *sigJ* impliqués dans la réponse au stress osmotique chez *S. coelicolor* (Viollier *et al.*, 2003a).

De plus, un autre niveau de compétition peut être révélé par la présence de promoteurs reconnus par le facteur  $\sigma$  WhiG en amont de la région codante du gène *basR*. Un de ces promoteurs chevauche en effet le promoteur *basRp1* et il est donc possible d'imaginer qu'une fois fixé sur son propre promoteur (fonctionnel ou non), le facteur  $\sigma$  WhiG empêcherait la reconnaissance et donc la transcription au site *basRp1* par  $\sigma^{\text{HasR}}$  lui-même.

La conservation de ce chevauchement entre le promoteur *sigBp1* du gène *sigB*, homologue des gènes *bas* chez *S. coelicolor*, et un promoteur putatif reconnu par le facteur  $\sigma$  WhiG rend cette hypothèse d'autant plus vraisemblable. Ce promoteur de type  $\sigma$  WhiG n'ayant cependant pas été décrit par les auteurs (Cho *et al.*, 2001), aucune indication sur sa potentielle fonctionnalité ne peut donc en être déduite.

### A-I.1.3 – MODULATION DE LA REGULATION CROISEE DES GENES HAS EN FONCTION DU STADE DE CROISSANCE

L'analyse de la transcription des gènes *bas* chez les mutants a permis de montrer que les facteurs  $\sigma^{\text{Has}}$  n'exercent pas le même type de régulation, c'est-à-dire une activation ou au contraire une inhibition de la transcription, selon la phase de croissance considérée. Cependant, que ce contrôle

soit positif ou au contraire négatif, il est important de noter qu'il pourrait révéler les mêmes types d'interactions, à savoir l'interaction d'un facteur  $\sigma^{\text{Has}}$  avec son propre promoteur ou l'interaction croisée d'un facteur  $\sigma^{\text{Has}}$  avec le promoteur de l'autre copie *bas*.

Nous pouvons donc représenter trois modèles de la régulation transcriptionnelle des gènes *bas*, un pour chaque phase de croissance (Figure D-1).

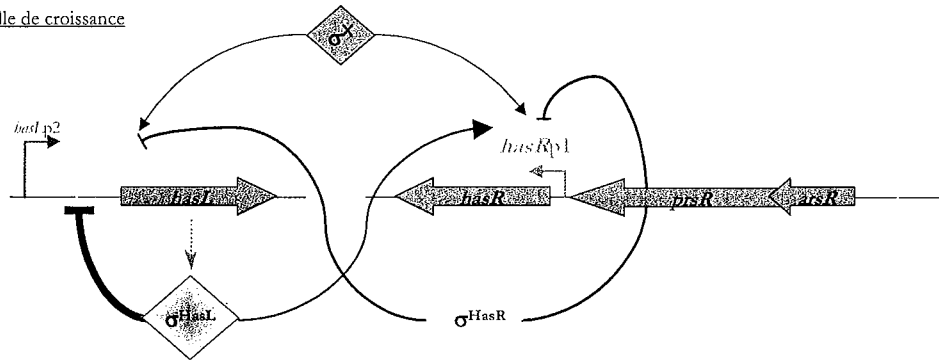
Dans ces modèles, nous ne considérerons pas les transcrits initiés aux sites *basRp2* et *basLp2* qui sont responsables de l'expression basale des gènes *bas* (et du système de régulation anti-anti- $\sigma$ /anti- $\sigma$  codé par les gènes *arsR* et *prsR* en amont de *basR*). De plus, ces modèles pourraient atteindre un niveau de complexité bien supérieur si nous intégrions les régulations de type post-transcriptionnel, notamment via les anti-facteurs  $\sigma$ , pouvant inhiber l'activité des facteurs  $\sigma^{\text{HasR}}$  et/ou  $\sigma^{\text{HasL}}$  (voir chapitre A-II).

Ainsi, le facteur  $\sigma^{\text{HasR}}$  est responsable de son autotranscription en phase de transition et en phase stationnaire de croissance, alors qu'il semble exercer un effet négatif sur son propre promoteur (*basRp1*) en phase exponentielle de croissance. Le contrôle du gène *basL* par le facteur  $\sigma^{\text{HasR}}$  est également phase-dépendant, puisqu'un phénomène d'inhibition transcriptionnelle est observé en phase exponentielle de croissance, alors qu'en phases de transition et stationnaire de croissance, le promoteur *basLp1* est positivement régulé par le facteur  $\sigma^{\text{HasR}}$ .

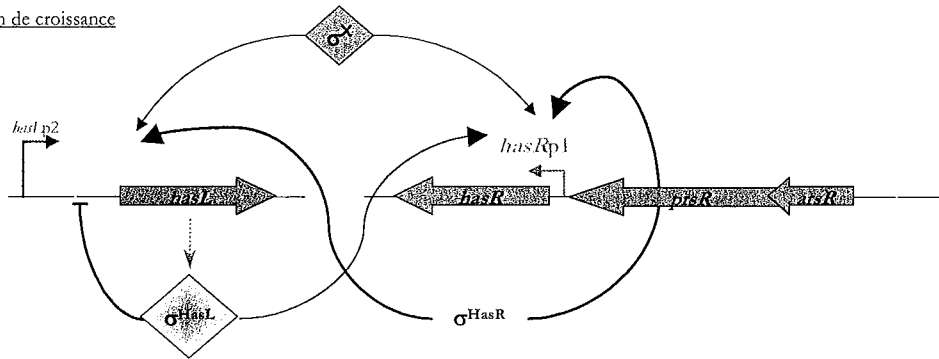
De façon similaire, la régulation exercée par le facteur  $\sigma^{\text{HasL}}$  est dépendante de la phase de croissance. En effet, le facteur  $\sigma^{\text{HasL}}$  inhibe son propre promoteur (*basLp1*) en phases exponentielle et de transition de croissance alors qu'il s'autoactive en phase stationnaire. Au contraire, le contrôle positif du gène *basR* (via *basRp1*) par le facteur  $\sigma^{\text{HasL}}$  est observé dans les trois phases de croissance.



A. Phase exponentielle de croissance



B. Phase de transition de croissance



C. Phase stationnaire de croissance

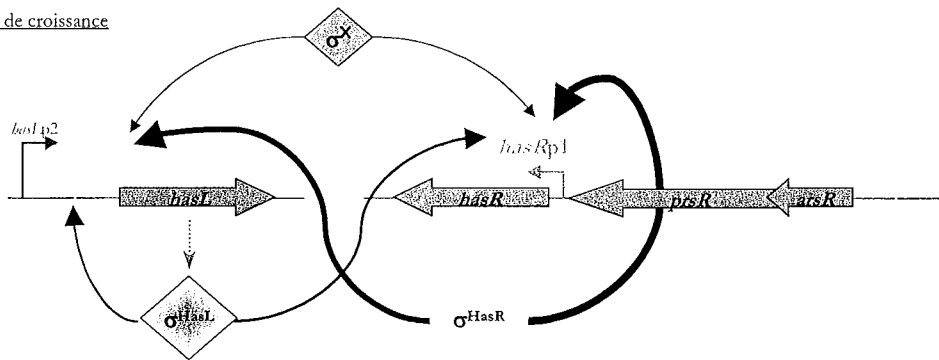


Figure D-1 : Modèles de régulation transcriptionnelle des gènes *has* en phase exponentielle (A.), de transition (B.) et stationnaire (C.) de croissance.

Seules les régulations des transcrits initiés aux sites *hasRp1* et *hasLp1* sont représentées. L'épaisseur des différentes flèches est proportionnelle à la force de la régulation exercée (estimée par PCR quantitative).

## A-1.2 $\propto$ L'ACTIVITE DES FACTEURS SIGMA HAS EST-ELLE REGULEE PAR DES MECANISMES POST-TRADUCTIONNELS ?

Les réponses adaptatives aux changements environnements et leurs systèmes de régulation correspondants observés chez les *Streptomyces* sont similaires à ceux décrits chez la bactérie du sol *Bacillus subtilis*. Chez cet organisme, différents facteurs  $\sigma$  assurent la coordination de l'expression des gènes en réponse aux conditions environnementales, aux niveaux métabolique et morphologique (voir paragraphe B-II.2.2 de l'introduction). Le facteur  $\sigma^{\text{B}}$  de réponse générale au stress ainsi que les facteurs  $\sigma$  de sporulation (la sporulation se mettant en place en réponse à un stress de carence nutritionnelle) sont sujets à divers niveaux de régulation post-traductionnelle.

### A-1.2.1 - SYNTHÈSE D'UN FACTEUR SIGMA SOUS FORME DE PROPOLYPEPTIDE (PRO-SIGMA)

La régulation spatio-temporelle de l'expression génique durant la différenciation des endospores chez *B. subtilis* implique l'activation séquentielle d'une cascade de facteurs  $\sigma$  (Kroos and Yu, 2000). La succession de l'activation de facteurs  $\sigma$  aboutit à la formation de deux compartiments cellulaires distincts, la cellule mère et l'endospore. Cette régulation au cours des stades de différenciation s'accompagne d'une régulation spécifique en fonction du compartiment cellulaire, avec établissement d'une "conversation intercellulaire croisée", c'est-à-dire la mise en place d'un contrôle concerté entre les compartiments. L'activité de ces différents facteurs  $\sigma$  est contrôlée à divers niveaux dont la synthèse sous forme inactive de facteur  $\sigma$ . L'activation résulte de la maturation par protéolyse de pro- $\sigma$  inactifs en  $\sigma$  actifs. Généralement, une vingtaine d'acides aminés de l'extrémité N-terminale sont ainsi clivés ; c'est le cas pour les facteurs  $\sigma^{\text{E}}$  et  $\sigma^{\text{K}}$  exprimés exclusivement dans la cellule mère (Haldenwang, 1995). Les protéases SpoIIGA et SpoIVFB sont impliquées dans le clivage des pro-séquences activant les facteurs  $\sigma^{\text{E}}$  et  $\sigma^{\text{K}}$  respectivement (Stragier *et al.*, 1988; Rudner *et al.*, 1999)

Chez *Streptomyces coelicolor*, le facteur  $\sigma^{\text{R}}$ , impliqué dans la réponse au stress oxydatif, est synthétisé sous forme d'un pro- $\sigma$  de 227 acides aminés, et gagne sa forme active en perdant les 4 premiers résidus de l'extrémité N-terminale (Paget *et al.*, 1998).

Le facteur  $\sigma^{\text{H}}$  de *S. coelicolor*, impliqué à la fois dans la différenciation et dans la réponse au stress (Kelemen *et al.*, 2001), est quant à lui retrouvé sous cinq isoformes différentes. Les trois premières, nommées SigH- $\sigma^{37}$ , SigH- $\sigma^{51}$  et SigH- $\sigma^{52}$ , sont le résultat de la présence de trois codons d'initiation de la traduction distincts (Viollier *et al.*, 2003a). De plus, les protéines SigH- $\sigma^{52}$  et SigH- $\sigma^{51}$  sembleraient être des pro- $\sigma$  de deux autres isoformes de 38 kDa et 34 kDa. De plus, ce mécanisme de régulation traductionnelle du facteur  $\sigma^{\text{H}}$  semble être impliqué dans la différenciation, puisque la protéolyse des facteurs SigH- $\sigma^{52}$  et SigH- $\sigma^{51}$  est associée à la formation du mycélium aérien.

Cependant, contrairement à *B. subtilis*, ces extensions N-terminales ne semblent pas être des inhibiteurs absolus de l'activité du facteur  $\sigma^{\text{H}}$  puisque les trois isoformes, SigH- $\sigma^{37}$ , SigH- $\sigma^{51}$  et SigH- $\sigma^{52}$ , présentent la même activité catalytique, une fois reconstituées avec le core enzyme de l'ARN polymérase. En revanche, aucun motif particulier, signature de la protéolyse, n'a pu être identifié et la nature de la protéase responsable de cette maturation n'est pas encore connue.

Aucun élément ne nous permet donc de suspecter une telle régulation pour les facteurs  $\sigma^{\text{Has}}$  et seules des analyses protéiques pourraient permettre de caractériser une telle maturation par protéolyse.

### **A-1.2.2 - LA PROTEOLYSE, MECANISME DE CONTROLE DE LA DIFFERENCIATION CELLULAIRE**

Le contrôle de l'activité de régulateurs centraux de la différenciation cellulaire fait également intervenir des mécanismes de dégradation protéolytique. Ce contrôle est spécifique d'un stade de développement donné ou d'une lignée cellulaire. Il constitue ainsi un mécanisme majeur de différenciation cellulaire par l'établissement de profils d'expression spécifiques (Gottesman, 1999).

Ainsi, les protéines présentant une succession d'acides aminés non polaires à leur extrémité C-terminale sont labiles, et rapidement dégradées par des protéases spécifiques. Cette extrémité non polaire peut être présente dans la protéine native ou être greffée par l'intermédiaire d'un ARN stable 10 S également appelé ARN SsrA codant un peptide de 10 acides aminés, aussi appelé "tag", dirigeant les protéines dont la traduction est interrompue vers la dégradation par des protéases telle que FtsH ou ClpAP (Gottesman *et al.*, 1998; Herman *et al.*, 1998). Ainsi, le contrôle de la durée de demi-vie des protéines régulatrices est utilisé dans la maintenance cellulaire (protéolyse des protéines anormales) mais aussi dans le contrôle de l'expression génique. Chez *E. coli*, le facteur  $\sigma^{32}$  régulant la réponse au choc thermique est la cible de la protéase FtsH (Herman *et al.*, 1995).

Plusieurs éléments permettent d'étayer l'hypothèse de la dégradation programmée de l'un des facteurs  $\sigma^{\text{Has}}$ . La séquence des produits de traduction putatifs des gènes *has* révèle que le produit HasL présente à son extrémité C-terminale la séquence Sérine-Alanine-Alanine alors que HasR présente la séquence Thréonine-Aspartate-Histidine. Ainsi, la protéine HasL pourrait être plus labile que HasR. Chez *S. coelicolor* A3(2), les gènes codant les sous-unités catalytiques (ClpP) et les sous-unités régulatrices (ClpX et ClpC) de la famille des protéases Clp ATP-dépendantes ont été identifiés (de Crécy-Lagard *et al.*, 1999; Viala *et al.*, 2000). Ces complexes protéasiques ont été impliqués dans la régulation de la différenciation morphologique et métabolique des *Streptomyces*. En effet, la fabrication d'un mutant nul (interruption au locus chromosomique) est associée à la déficience de la différenciation du mycélium aérien. En revanche, la surexpression des gènes accélère la différenciation et accroît la production d'antibiotiques (de Crécy-Lagard *et al.*, 1999).

### A-I.2.3 - CONTROLE DE L'ACTIVITE PAR DES INHIBITEURS (ANTI-FACTEUR SIGMA)

Le facteur  $\sigma$  WhiG, acteur de la différenciation tardive chez les *Streptomyces*, présente une transcription constitutive à partir d'un site unique d'initiation de la transcription (Kelemen *et al.*, 1996). Impliqué dans la formation des spores,  $\sigma$  WhiG présente une activité durant les phases tardives. Des données suggèrent fortement que l'activité de  $\sigma$  WhiG est régulée par un anti-facteur  $\sigma$ . En effet,  $\sigma$  WhiG complémente l'action de son  $\sigma$  homologue chez *S. typhimurium*,  $\sigma^{\text{FlhA}}$ , impliqué dans la synthèse des flagelles. Un contrôle négatif de l'activité de  $\sigma$  WhiG a été mis en évidence par l'action du produit du gène *flgM* codant l'anti- $\sigma^{\text{FlhA}}$  (Kelemen *et al.*, 1996).

Le facteur sigma  $\sigma^{\text{R}}$ , impliqué dans la réponse au stress oxydatif (Paget *et al.*, 1998), est également régulé par anti-facteur  $\sigma$ , nommé RsrA, dont le gène est localisé immédiatement en aval de *sigR* (Kang *et al.*, 1999). Dans des conditions non stressantes, RsrA, dans son état réduit, se lie au facteur  $\sigma^{\text{R}}$  et l'empêche de transcrire les gènes qui sont sous son contrôle. En revanche, lors d'un stress oxydatif, un ou plusieurs ponts disulfures se forment dans la protéine RsrA, qui perd alors son affinité pour le facteur  $\sigma^{\text{R}}$ . Ce dernier est alors libre de se fixer à l'ARN polymérase et d'activer la transcription de ses gènes cibles, dont *trxB* et son propre gène *sigR*. L'expression de l'opéron *trxB* codant le système thiorédoxine/thiorédoxine réductase, permet d'une part un retour à un niveau normal de réduction des groupements thiols et d'autre part la réduction de RsrA, ce qui permet de fermer la boucle de régulation (Kang *et al.*, 1999, Figure D-2).

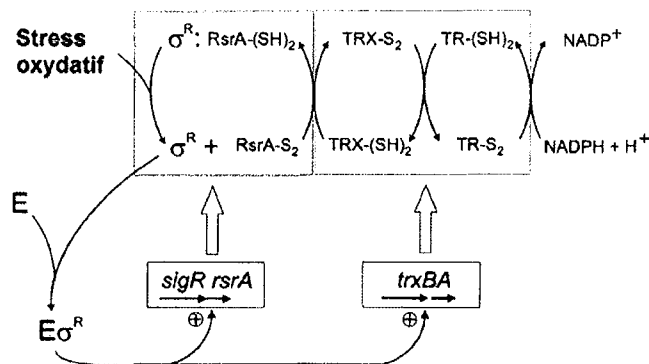


Figure D-2 : Mode de régulation du facteur  $\sigma^{\text{R}}$  en réponse à un stress oxydatif chez *S. coelicolor*. Selon (Kang *et al.*, 1999).

La régulation du facteur  $\sigma^{\text{H}}$  par un anti-facteur  $\sigma$ , nommé PrsH, de *S. coelicolor* démontre également l'importance de cette régulation par les anti-facteur  $\sigma$  chez les *Streptomyces* (Sevcikova and Kormanec, 2002; Viollier *et al.*, 2003b). De plus, l'interruption du gène *prsH* dans un contexte mutant pour le gène *sigH* (où le mutant *sigH* seul n'a pas de phénotype) entraîne le blocage de la

différenciation au stade du mycélium végétatif (Viollier *et al.*, 2003a), suggérant que cet anti-facteur  $\sigma$  est nécessaire pour l'inhibition d'autres régulateurs.

Cette hypothèse est étayée par le phénotype d'altération de la sporulation conféré par l'interruption d'autres anti-facteur  $\sigma$  chez les *Streptomyces* et notamment de RsrA, l'anti-facteur  $\sigma^{\text{R}}$  (Paget *et al.*, 2001a) et de RsuA, l'anti facteur  $\sigma$  du facteur  $\sigma$  codé par le gène *sigU* (Gehring *et al.*, 2001).

Enfin, il est intéressant de noter que parmi la famille des neuf facteurs  $\sigma$  homologues du facteur  $\sigma^{\text{B}}$  de *B. subtilis*, deux des gènes correspondant sont adjacents à des gènes codant des anti-anti facteur  $\sigma$  et des anti-facteur  $\sigma$  (*sigI* et *sigJ*) et trois sont adjacents à seulement un gène codant un anti-facteur  $\sigma$  (*sigH*, *sigL* et *sigM*). Une étude récente sur la régulation post-traductionnelle  $\sigma^{\text{H}}$  a cependant montré que, bien que les protéines SigJ et SigI « cross-réagissaient » avec l'anticorps spécifique du facteur  $\sigma^{\text{H}}$ , elles n'interagissaient pas avec l'anti-facteur  $\sigma$  PrsH (Viollier *et al.*, 2003b).

L'analyse de l'environnement génétique des gènes *has* a permis de révéler la présence de deux gènes codant des anti-facteurs  $\sigma$  (*prsR* et *prsL*) à proximité immédiate des gènes *hasR* et *hasL*.

Le modèle de régulation des gènes *has* atteint donc un niveau de complexité bien supérieur, une fois ces niveaux de régulation post-traductionnels pris en compte (Figure D-3). L'organisation opéronique du gène *hasR* avec le système anti- $\sigma$ /anti-anti- $\sigma$  codé par les gènes *prsR* et *arsR* respectivement suggère un contrôle du facteur  $\sigma^{\text{HasR}}$  par ce système. Il est cependant également envisageable que l'anti-facteur  $\sigma$  codé par le gène *prsR* régule le facteur  $\sigma^{\text{HasL}}$ . De la même façon, l'anti-facteur  $\sigma$  codé par le gène *prsL*, en aval du gène *hasL*, pourrait aussi bien réguler l'activité du facteur  $\sigma^{\text{HasL}}$  que celle du facteur  $\sigma^{\text{HasR}}$ . De plus, ces anti-facteur  $\sigma$  pourraient ne pas présenter la même affinité pour chacun des deux facteurs  $\sigma^{\text{HasR}}$  et  $\sigma^{\text{HasL}}$  (voir paragraphe Perspectives).

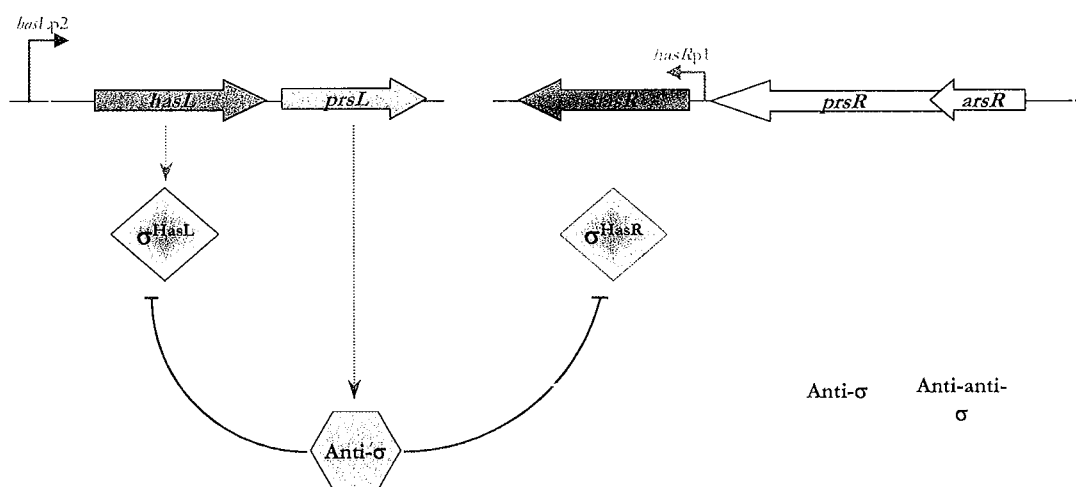


Figure D-3 : Modèle de régulation de l'activité des facteurs  $\sigma^{\text{Has}}$  faisant intervenir les systèmes anti-anti- $\sigma$  / anti- $\sigma$ .

- CHAPITRE A-II -

LA REPONSE AU STRESS CHEZ LES  
*STREPTOMYCES*

## A-II.1 $\oslash$ LA REDONDANCE DES SYSTEMES DE REGULATION TRANSCRIPTIONNELLE

Lors de ce travail, l'analyse phénotypique des mutants interrompus pour le(s) gène(s) *has* n'a révélé qu'une faible sensibilité des mutants *has* au stress oxydatif. Ce phénotype a été confirmé par l'induction transcriptionnelle des promoteurs *hasRp1* et *hasLp1* suite à un stress oxydatif ou osmotique. Il a également été montré que cette induction n'était pas exclusivement dépendante des facteurs  $\sigma^{\text{Has}}$  mais au contraire faisait intervenir d'autres facteurs  $\sigma$  de type  $\sigma^{\text{B}}$ . De plus, l'analyse transcriptionnelle des gènes *has* dans les contextes mutants a également révélé l'intervention d'autres facteurs  $\sigma$ , reconnaissant donc le même type de séquences promotrices, dans la régulation de leur expression. L'ensemble de ces données révèle la probable redondance des facteurs  $\sigma$  impliqués dans la réponse au stress chez *Streptomyces*. Parmi les 9 gènes codant les facteurs  $\sigma$  de type  $\sigma^{\text{B}}$  révélés par l'analyse de la séquence du génome de *S. coelicolor* (Bentley *et al.*, 2002), quatre ont été isolés et caractérisés ( $\sigma^{\text{F}}$ ,  $\sigma^{\text{G}}$ ,  $\sigma^{\text{H}}$  et SigB). Le dernier, également appelé  $\sigma^{\text{J}}$ , est l'orthologue des facteurs  $\sigma^{\text{Has}}$ , et notamment de  $\sigma^{\text{HasR}}$  avec une conservation de l'environnement génétique. Cependant les résultats obtenus au cours de ce travail sont différents de ceux obtenus par (Cho *et al.*, 2001) sur le facteur  $\sigma$  SigB chez *S. coelicolor*. En effet, le mutant *sigB* de *S. coelicolor* est affecté dans la différenciation, puisqu'il apparaît comme étant déficient dans la formation du mycélium aérien (mutant *bald*), et également dans la défense contre le stress osmotique (sensibilité à l'hyperosmolarité). Cette différence phénotypique peut notamment révéler un contrôle transcriptionnel différent entre les gènes *has* chez *S. ambifaciens* et leur orthologue *sigB* chez *S. coelicolor*, mais peut également être expliquée par la présence d'un facteur  $\sigma$  de type  $\sigma^{\text{B}}$  supplémentaire chez *S. ambifaciens*, de par la duplication des gènes *has* (voir paragraphe B-II.2).

Selon Viollier *et al.* (2003a), la réponse au stress osmotique chez *S. coelicolor* dépendrait d'un système senseur qui coordonnerait l'activité de multiples facteurs  $\sigma$  paralogues pour le contrôle de l'expression des gènes d'un même régulon de stress. Le facteur  $\sigma^{\text{J}}$ , orthologue des facteurs  $\sigma^{\text{Has}}$  (appelé SigB par Cho *et al.* (2001)) appartiendrait à ce groupe de facteurs  $\sigma$  paralogues impliqués dans la réponse au stress osmotique, puisque la protéine SigJ apparaît comme étant accumulée dans des conditions de stress osmotique.

Ainsi, les systèmes de réponse au stress osmotique chez les *Streptomyces* dépendraient de l'activité de plusieurs facteurs  $\sigma$  spécifiques, expliquant l'absence de phénotype marqué des mutants *has* chez *S. ambifaciens*.

**A-II.2  $\sigma$  LES FACTEURS SIGMA : UN LIEN ENTRE DIFFÉRENCIATION MORPHOLOGIQUE ET RÉPONSE AU STRESS**

Le chevauchement entre le promoteur putatif reconnu par le facteur  $\sigma$  WhiG et le promoteur *basRp1* en amont du gène *basR* permet d'émettre l'hypothèse d'un lien entre la différenciation morphologique, régulée par le facteur  $\sigma$  WhiG chez *S. coelicolor* (Chater *et al.*, 1989), et la réponse au stress.

Cette hypothèse a également été formulée lors de la caractérisation du facteur  $\sigma^H$  chez *S. coelicolor* (Kelemen *et al.*, 2001; Sevcikova *et al.*, 2001). La première équipe a en effet montré que le facteur  $\sigma^H$  était à la fois impliqué dans la différenciation morphologique et dans la réponse au stress osmotique puisque le mutant interrompu pour le gène *sigH* était affecté dans sa croissance en conditions de stress salin mais également dans sa sporulation (Sevcikova *et al.*, 2001). Les travaux de Kelemen *et al.* (2001) ont par ailleurs montré que la transcription du gène *sigH* initiée au promoteur *sigHp2*, inductible en conditions de stress (Kormanec *et al.*, 2000), était régulée par le répresseur transcriptionnel BldD, nécessaire à la formation des hyphes aériens chez *S. coelicolor* (Willey *et al.*, 1993).

Enfin, une étude statistique beaucoup plus globale, utilisant des approches de protéomiques, a montré le contrôle au cours du développement des régulons de stress chez *S. coelicolor* (Van Deutekom *et al.*, 1996).



B - SECONDE PARTIE :

CREATION DE DIVERSITE  
GENETIQUE PAR REARRANGEMENT  
D'ADN

**- CHAPITRE B-I -**

**POLYMORPHISME DES EXTREMITES  
DES REPLICONS LINEAIRES**

La circularité du chromosome a été décrite chez de nombreuses espèces bactériennes, la linéarité des chromosomes étant jusqu'à peu l'apanage des eucaryotes. Cependant, la linéarité chromosomique a également été observée chez certains genres et espèces bactériens tels que les *Streptomyces*. La distribution et la diversité des structures télomériques montrent que la linéarité a été acquise de façon sporadique chez les bactéries. Cependant, un fort polymorphisme intraspécifique et interspécifique caractérise les régions subtélomériques des chromosomes des organismes possédant un (ou des) chromosome(s) linéaire(s).

## B-I.1 $\infty$ DISTRIBUTION DE LA LINEARITE CHROMOSOMIQUE CHEZ LES BACTERIES

La première mise en évidence de la linéarité chromosomique chez une bactérie date de 1989, lorsque Ferdows et Barbour montrèrent que *Borrelia burgdorferi* possède un chromosome linéaire (Ferdows and Barbour, 1989). Puis en 1993, Lin *et al.* prouvèrent la linéarité de l'ADN chromosomique de *Streptomyces lividans* (Lin *et al.*, 1993). Elle semble être répandue à l'ensemble du genre *Streptomyces* puisque *S. coelicolor*, *S. ambofaciens*, *S. griseus*, *S. rimosus*, *S. avermitilis* et *S. hygroscopicus* possèdent également un ADN chromosomique linéaire (Lezhava *et al.*, 1995b; Leblond *et al.*, 1996; Redenbach *et al.*, 1996; Pandza *et al.*, 1997; Omura *et al.*, 2001; Pang *et al.*, 2002). La linéarité a aussi été démontrée chez *Saccharopolyspora erythraea* (anciennement *Streptomyces erythreus*, (Reeves *et al.*, 1998) et suggérée chez *Actinoplanes*, *Amycolatopsis*, *Micromonospora*, *Nocardia*, *Rhodococcus* et *Streptoverticillium* (Redenbach *et al.*, 2000). Tous ces genres bactériens appartiennent à l'ordre des Actinomycétales. Cependant, toutes les espèces bactériennes de cet ordre ne possèdent pas un chromosome linéaire. En effet, l'ADN chromosomique des Mycobactéries et des Corynébactéries, notamment, est circulaire (Fu and Fu-Liu, 2002; Tauch *et al.*, 2002).

Ainsi, la distribution de la linéarité chromosomique dans un arbre phylogénétique d'espèces procaryotes (Volf and Altenbuchner, 2000 ; Figure D-4) souligne le caractère dispersé de la linéarité chromosomique. En effet, elle est présente non seulement chez les bactéries Gram positives à haut pourcentage en bases G et C (Actinomycétales) mais aussi chez les Spirochètes (*Borrelia burgdorferi*) et les Protéobactéries  $\alpha$  (*Agrobacterium tumefaciens*) (Wood *et al.*, 2001). De plus, la linéarité du chromosome a été suggérée chez une Protéobactérie  $\gamma$  (*Coxiella burnetii*) (Willems *et al.*, 1998).

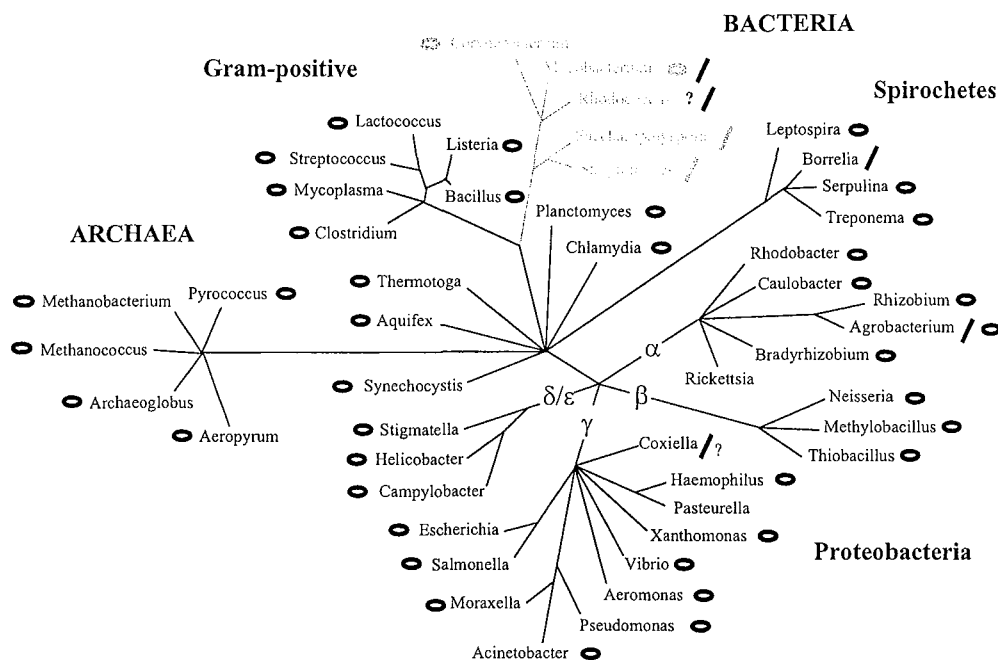


Figure D-4 : Distribution de la linéarité chromosomique sur un arbre phylogénétique des espèces procaryotes selon Volf *et al.* (2000)

Un anneau noir correspond à la présence d'un chromosome circulaire chez l'espèce considérée, un trait noir à un chromosome linéaire, un trait rouge à un plasmide linéaire. Le point d'interrogation signifie que le chromosome a été montré comme étant à la fois circulaire et linéaire. L'ordre des Actinomycétales est représenté en gris.

Il semble que la distribution de la linéarité plasmidique soit plus large que celle des chromosomes. Ainsi, des plasmides linéaires peuvent coexister avec un chromosome linéaire et des plasmides circulaires, par exemple chez les *Borrelia* et les *Streptomyces*.

De plus, la présence de plasmides linéaires a été montrée dans des genres bactériens possédant des chromosomes circulaires, comme *Mycobacterium* (Picardeau and Vincet, 1997), *Nocardia* (Kalkus *et al.*, 1990) et *Rhodococcus* (Kalkus *et al.*, 1993).

Des plasmides linéaires ont aussi été retrouvés chez *Clavibacter michiganensis* (Brown *et al.*, 2002) et *Planobispora rosea* (Polo *et al.*, 1998). Il est à noter que des plasmides circulaires ont été observés chez toutes ces Actinomycétales hormis *P. rosea*.

Au sein des Protéobactéries, des plasmides linéaires ont été mis en évidence chez *Xanthobacter* (subdivision  $\alpha$ ) (Bergeron *et al.*, 1998; Krum and Ensign, 2001) et chez *Klebsiella oxytoca* (subdivision  $\gamma$ ) (Li and Lim, 2000). Chez cette dernière, des plasmides circulaires ont également été détectés.

La linéarité chromosomique chez les bactéries serait donc issue d'événements sporadiques. De plus, la linéarité semble être restreinte aux Spirochètes, aux Actinomycétales et aux Protéobactéries (hormis les subdivisions  $\delta$  et  $\varepsilon$ ) à la fois pour les plasmides et les chromosomes.

## B-I.2 $\propto$ REPLICONS LINEAIRES ET REARRANGEMENTS CHROMOSOMIQUES

Les bornes des réarrangements chromosomiques sont localisées dans les régions subtélomériques (voir paragraphe A-II.2.3 de l'introduction). La question traitée dans ce paragraphe concerne le rôle de la linéarité chromosomique dans l'instabilité génomique. Seront ensuite exposés et discutés les mécanismes moléculaires impliqués dans la formation des réarrangements chromosomiques et enfin, l'impact de ces réarrangements sur l'architecture des génomes.

### B-I.2.1 – LA LINEARITE CHROMOSOMIQUE CONFERE-T-ELLE UN AVANTAGE ?

#### B-I.2.1.1 – REGIOSPECIFICITE DES REARRANGEMENTS CHEZ LES STREPTOMYCES

L'instabilité génétique chez les *Streptomyces* a été corrélée de longue date à la formation de réarrangements chromosomiques de grande ampleur. Deux grands types de réarrangements ont été caractérisés : délétion de grande taille et amplification de motif d'ADN ou "Amplifiable Unit of DNA" (AUD) (pour revue, Leblond and Decaris, 1999). La région touchée par les réarrangements, ou région instable, non essentielle dans les conditions de laboratoire, représente environ 2 Mb soit un quart du génome chez *S. ambofaciens* DSM40697.

Depuis la mise en évidence de la linéarité de l'ADN chromosomique chez les *Streptomyces* (en 1993 par Lin *et al.* (1993) chez *S. lividans*), région instable et régions terminales du chromosome se superposent. En effet, la quasi totalité des réarrangements décrits se déroule aux extrémités du chromosome.

La séquence complète du génome de *S. coelicolor* A3(2) (Bentley *et al.*, 2002) nous indique que les régions subtélomériques sont riches en gènes dits de «contingence» (voir paragraphe I.1.3). Cette compartimentation expliquerait la tolérance pour les réarrangements dans ces régions du chromosome. L'instabilité accrue de ces régions par rapport au reste du chromosome pourrait avoir façonné l'organisation génétique et exclu les gènes « essentiels » des régions subterminales.

Par ailleurs, le « core » chromosomique, correspondant à la région centrale du chromosome et contenant les gènes essentiels et notamment l'origine de réplication, semble peu permissif aux réarrangements. En effet, l'analyse du polymorphisme des régions ITS chez deux souches de *S. ambofaciens* (DSM40697 et ATCC23877, Wenner *et al.*, 2002) a montré que, bien que la recombinaison homologue participe fortement à la création de variabilité de ces loci, aucun réarrangement de grande ampleur, contrairement à d'autres systèmes bactériens, n'a pu être

déte t  entre les organisations chromosomiques des deux souches. En effet, la recombinaison homologue entre loci *rrm* entraîne l' change de blocs de s quences sans  change des r gions flanquantes (crossing-over).

Ainsi, les r arrangements chromosomiques sont confin s aux extr mit s du chromosome lin aire des *Streptomyces*.

### B-I.2.1.2 – LA LINEARITE CHROMOSOMIQUE FAVORISE-T-ELLE L'INSTABILITE CHROMOSOMIQUE ?

Plusieurs propri t s intrins ques   la structure lin aire du chromosome pourraient favoriser l'apparition de r arrangements chromosomiques dans cette r gion :

Les chromosomes des *Streptomyces* poss dent une origine de r plication unique localis e approximativement    quidistance des extr mit s chromosomiques. Cette origine fonctionne sur un mode bidirectionnel en direction des extr mit s (Musialowski *et al.*, 1994; Fischer *et al.*, 1998a). Il est probable, compte tenu du mode pr sum  d'acquisition de la lin arit  chromosomique (voir ci-apr s, paragraphe B-I.2.3.1.3 de la discussion), que des sites de terminaison de la r plication (sites Ter chez *E. coli*) soient dispers s dans les r gions subterminales. A d faut, les t lom res constitueraient des termini obligatoires. Chez *E. coli*, la zone du terminus de la r plication est une r gion hautement recombinog ne. L' quipe de B. Michel et S.D. Ehrlich a montr  que l'augmentation du taux de recombinaison r sultait de l'apparition de cassures double brin g n r e par l'arr t de la fourche de r plication aux sites de pause (Bidnenko *et al.*, 2002).

La seule pr sence d'extr mit s d'ADN peut  galement accro tre les fr quences de recombinaison impliquant des s quences subt lom riques. En effet, une d ficience dans les fonctions de maintien des t lom res, notamment chez la levure, est associ e   la formation de grands r arrangements chromosomiques (Hackett *et al.*, 2001) analogues   ceux retrouv s chez les *Streptomyces*.

Enfin, les r gions subt lom riques sont riches en transposases (Bentley *et al.*, 2002). Le seul cas d'implication directe de la transposition d' l ments mobiles dans les r arrangements d'ADN correspond   l'amplification d'ADN chez *S. lividans* lors d'exp riences d'interruption de s quences dans les r gions t lom riques (Gunes *et al.*, 1999). En effet, la transposition g n re des cassures dans l'ADN (aux bornes des  l ments ainsi qu'aux sites cibles) qui sont susceptibles d'initier des  v nements de recombinaison.

Cependant, outre l'activit  endonucl asique de la transposase, les nombreuses copies d'IS et autres transposons sont susceptibles de fournir un substrat   la recombinaison homologue et ainsi de participer indirectement   la formation de grands r arrangements.

La forte concentration d' l ments mobiles dans les r gions terminales peut donc   ces deux titres accro tre les fr quences de recombinaison dans ces r gions chromosomiques.

Enfin, lors de l'analyse des séquences impliquées aux bornes de certains réarrangements chez *S. glaucescens*, il a été noté la présence de séquences palindromiques ou de répétitions directes imparfaites à proximité des jonctions. Ces séquences présentent des homologies avec le consensus de reconnaissance de la recombinase site-spécifique du plasmide intégratif pIJ408. Les auteurs ont alors suggéré que les réarrangements pourraient découler de la reconnaissance erronée de certains sites par l'intégrase d'un plasmide intégratif (Birch *et al.*, 1991).

Le décrochage de la fourche de réplication, la déficience du fonctionnement télomérique, l'interaction avec une transposase ou une intégrase site-spécifique ont pour point commun la production de cassures double brin dans l'ADN. La réparation erronée de telles cassures pourrait être à l'origine des réarrangements.

### **B-1.2.2 - MECANISMES MOLECULAIRES DES RÉARRANGEMENTS CHROMOSOMIQUES**

Deux mécanismes moléculaires principaux de réparation des cassures double brin ont été mis en évidence chez les eucaryotes et les procaryotes : la recombinaison illégitime et la recombinaison homologue. Elles se distinguent par la taille de l'homologie des séquences impliquées dans les réarrangements.

La recombinaison illégitime a été impliquée à de nombreuses reprises dans les réarrangements chromosomiques chez les *Streptomyces*, qu'ils soient spontanés ou induits :

Chez *S. rimosus* par exemple, où un échange d'extrémités entre le chromosome et un plasmide linéaire a été caractérisé, une microhomologie de 4 pb a été identifiée au point de recombinaison (Pandza *et al.*, 1998).

De même, chez le mutant *S. ambofaciens* NSA27, la fusion chromosomique fait intervenir une microhomologie de 6 pb.

Enfin, lorsque les chromosomes ou plasmides sont sauvegardés sous forme circulaire chez des souches déficientes pour le maintien des télomères (mutation *trans* de la protéine Tpg ou *ais* dans les palindromes télomériques), le séquençage des bornes des réarrangements implique la recombinaison illégitime dans le mécanisme de réparation (Qin and Cohen, 2002).

Des événements de recombinaison homologue entre gènes dupliqués ont par ailleurs permis d'expliquer la structure chromosomique de souches mutantes de *Streptomyces* :

Les remplacements d'extrémités chromosomiques mis en évidence dans les souches NSA135HPP et NSA855, résultent d'événements de recombinaison homologue entre les gènes *hasR* et *hasL* (Fischer *et al.*, 1998b). La fréquence d'un tel événement ne peut être déduite des expériences réalisées.

Le remplacement de bras chromosomique a également été décrit chez un mutant de *S. griseus* issu d'une mutagenèse aux ultraviolets (Zenitani *et al.*, 1999). Le réarrangement implique un événement de recombinaison homologue entre deux copies d'un gène, codant des lipoprotéines

putatives et présentant plus de 99% d'identité en nucléotides (ORF-L 654 nt, ORF-R 651 nt). L'homologie de séquence cesse en 5' et 3' des deux ORF. Les deux copies ORF-R et ORF-L sont localisées à l'extérieur des TIR respectivement à 250 kb et 450 kb des extrémités des bras droit et gauche du chromosome linéaire de *S. griseus*. Le remplacement de bras aboutit à la duplication de 450 kb d'ADN dérivant du bras gauche à l'extrémité du bras droit délété, et à l'accroissement de la taille des TIR de 24 kb chez la souche sauvage (Lezhava *et al.*, 1995a) à 450 kb chez la souche mutante.

Ce cas est donc en tout point analogue aux remplacements de bras chromosomiques observés chez *S. ambofaciens*.

Ainsi, les événements de recombinaison homologue pourraient être fréquents à l'intérieur des répétitions terminales inversées (TIR). Compte tenu de l'orientation des répétitions terminales, ces événements mènent au remplacement des extrémités chromosomiques. Ces événements passeraient inaperçus puisque (i) non associés à des délétions géniques et (ii) conduisant au remplacement à l'identique de l'extrémité chromosomique. Ces événements conduiraient à homogénéiser les répétitions terminales. Cette hypothèse est étayée par l'absence quasi totale de polymorphisme entre les 210 kb constituant les TIR chez *S. ambofaciens* (Leblond *et al.*, 1996).

### **B-I.2.3 - REARRANGEMENTS CHROMOSOMIQUES ET EVOLUTION RAPIDE DES GENOMES**

L'étude des réarrangements subtélomériques au niveau intraclonal, c'est-à-dire en cultures pures dans les conditions de laboratoire, permet de visualiser l'impact de la recombinaison terminale sur les potentialités d'évolution rapide de l'information existante : formation de grandes délétions et duplications, amplification en tandem et formation de gènes chimériques aux points de recombinaison. En revanche, la comparaison des régions terminales au niveau intraspécifique ou interspécifique, de même que l'analyse de la séquence génomique de *S. coelicolor* A3(2) (Bentley *et al.*, 2002) et de *S. avermitilis* (Ikeda *et al.*, 2003), permettent, entre autres, d'estimer l'impact de cette plasticité génétique dans l'acquisition et l'échange d'information issue du transfert horizontal.

La question de l'impact de tels réarrangements sur l'évolution de la structure chromosomique est également soulevée par la présence de répétitions de séquences dans les régions terminales du chromosome, comme l'illustre la duplication des gènes *bas* chez *S. ambofaciens*. En effet, la présence de séquences répétées dans les régions terminales du chromosome linéaire des *Streptomyces* favorise l'échange d'extrémités chromosomiques. Ce type d'événement peut avoir deux conséquences majeures : la duplication de régions génomiques lorsque les échanges sont intraspécifiques, et le transfert horizontal d'information génétique lorsque les échanges sont interspécifiques.



### B-I.2.3.1 - RECOMBINAISON SUBTELOMERIQUE ET ACQUISITION D'INFORMATION GENETIQUE

#### B-I.2.3.1.1 - Transfert horizontal d'information génétique chez les *Streptomyces*

De nombreux résultats permettent d'appréhender l'importance du transfert horizontal sur l'évolution du génome des *Streptomyces*, et notamment l'impact de la recombinaison terminale sur l'acquisition et l'échange de séquences.

Tout d'abord, la comparaison des séquences composant les TIR, entre souches appartenant à l'espèce *S. ambofaciens* (DSM40697 et ATCC23877) révèle que des régions spécifiques de souche existent aux extrémités chromosomiques (Borges, 2001; Gallois, 2003). Ainsi, les deux souches DSM40697 et ATCC23877 présentent respectivement 50,7 kb et 33,2 kb d'ADN spécifique de souche (Figure D-5 ; voir ci-après, paragraphe B-I.2.3.1.2).

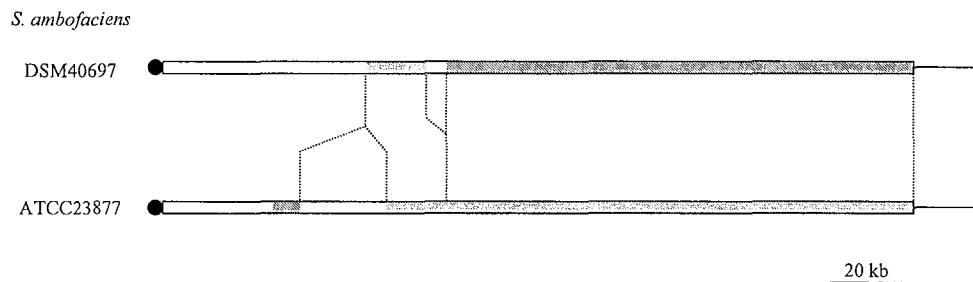


Figure D-5 : Comparaison des répétitions terminales inversées (rectangles) de deux souches de *Streptomyces ambofaciens*, DSM40697 (en haut) et ATCC23877 (en bas).

Les régions homologues entre les deux répétitions sont indiquées en zone hachurée, les régions spécifiques à chaque souche par des zones blanches. Les traits en pointillés indiquent les interruptions entre les zones homologues et les cercles noirs représentent les protéines terminales (Tpg).

Au niveau interspécifique, le même type de donnée ressort de l'hybridation des séquences couvrant les TIR de *S. lividans* sur l'ADN génomique de *S. coelicolor* A3(2) puisque aucune homologie n'a pu être détectée sur plus de 40 kb (Huang *et al.*, 1998).

La cartographie comparée d'espèces proches telles que *S. coelicolor* A3(2) et *S. ambofaciens* DSM40697, nous permet également de visualiser le polymorphisme des régions terminales chez les *Streptomyces*. Alors que la cartographie d'une vingtaine de marqueurs génétiques dans la région «core» du chromosome de *S. ambofaciens* DSM40697 et de *S. coelicolor* A3(2) montre une bonne conservation de l'ordre et de la localisation approximative de ces loci (Leblond *et al.*, 1996), Figure R-22), les régions terminales présentent une organisation différente.

De nombreux exemples étayent cette conclusion. Par exemple, la localisation du locus *hasR* (anti-anti sigma, anti sigma et sigma : *arsR*, *prsR* et *hasR*), sur le bras chromosomique droit de *S. ambofaciens* DSM40697, alors que son homologue est retrouvé sur le bras chromosomique

gauche de *S. coelicolor* A3(2) (*rsbB*, *rsbA* et *sigB* ou *sig8*, (Cho *et al.*, 2001)), distingue les deux espèces aux liens phylogénétiques pourtant étroits (Takeuchi *et al.*, 1996; Wenner *et al.*, 2002).

De plus, le locus chromosomique du gène *basL* n'est pas présent dans le génome de *S. coelicolor* A3(2). Enfin, l'hybridation des cosmides recouvrant environ 150 kb incluant le locus *basR* révèle des séquences localisées sur le bras gauche de *S. coelicolor* A3(2). Ce résultat indique qu'un réarrangement de grande ampleur s'est produit chez l'une ou l'autre des espèces depuis leur séparation à partir de leur dernier ancêtre commun.

Le séquençage des TIR de la souche *S. ambofaciens* ATCC23877 (projet Génoscope-IGM Orsay-LGM Nancy) permet de comparer beaucoup plus finement l'organisation génétique de *S. ambofaciens*, *S. coelicolor* A3(2) et *S. avermitilis*.

Les protéines potentielles déduites des ORF décelées sont comparées à celles des bases de données, et notamment celles de *S. coelicolor* (ScoDB) au Centre Sanger et de *S. avermitilis* (<http://avermitilis.ls.kitasato-u.ac.jp/>).

La comparaison révèle que :

□ une faible proportion d'ORF de *S. ambofaciens* présente un homologue «fort» (>75% d'identité en acides aminés) dans le génome de *S. coelicolor* ou de *S. avermitilis*.

En effet, à peine plus de la moitié des protéines déduites des séquences des TIR de *S. ambofaciens* ATCC22877 a un homologue codé par les génomes de *S. coelicolor* ou *S. avermitilis* (Figure D-6).

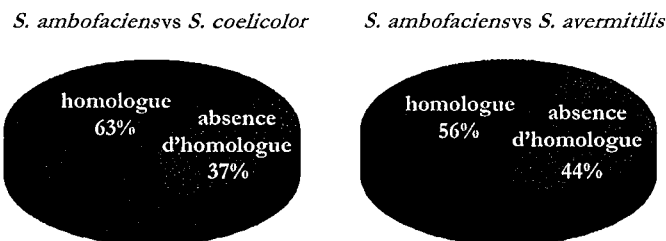


Figure D-6 : Pourcentages de protéines homologues entre les répétitions terminales inversées de *Streptomyces ambofaciens* ATCC23877 et les génomes de *S. coelicolor* et *S. avermitilis*.

Le terme homologue est défini selon les paramètres suivants : 30% d'identité protéique avec un chevauchement minimum de 80% de la longueur des protéines.

De plus, lorsqu'une protéine trouve son homologue chez *S. coelicolor* ou *S. avermitilis*, les pourcentages d'identité protéique entre les deux protéines putatives sont faibles (inférieurs à 50% pour plus de la moitié des protéines, Figure D-7).

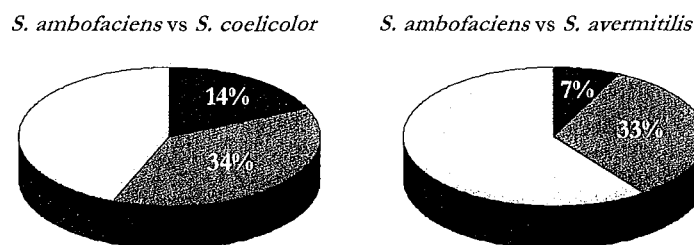
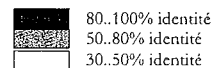


Figure D-7 : Pourcentages d'identité protéique des protéines homologues entre les répétitions terminales inversées de *Streptomyces ambofaciens* ATCC23877 et les génomes de *S. coelicolor* et *S. avermitilis*.

Les codes couleurs représentent les différents niveaux d'identité protéique :



□ lorsqu'un homologue existe, il est localisé à un autre locus et le niveau de syntenie est très faible.

En effet, les homologues des gènes localisés dans les TIR, donc aux extrémités, du chromosome de *S. ambofaciens* ATCC23877, sont répartis sur toute la longueur du génome de *S. coelicolor* (Figure D-8).

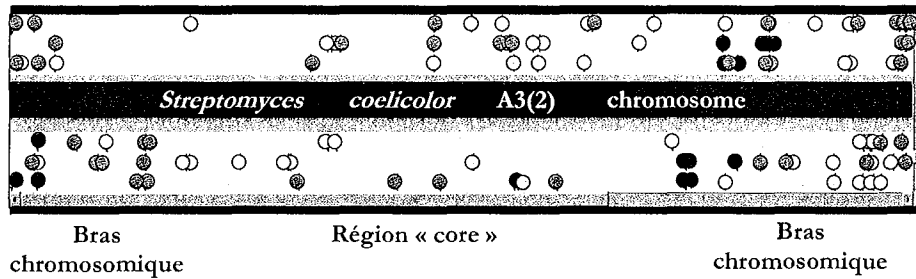


Figure D-8 : Localisation des homologues des gènes des TIRs de *S. ambofaciens* sur le chromosome de *S. coelicolor*.

Chaque point représente une protéine de *S. coelicolor*, homologue des produits des gènes identifiés dans les TIR de *S. ambofaciens* ATCC23877. Le symbolisme de couleur utilisé pour les identités protéiques est identique à celui de la Figure D-7.

Il est cependant important de rappeler que ces comparaisons se limitent aux régions terminales du chromosome de *S. ambofaciens*. Ainsi, il est probable que des homologues forts des protéines de *S. coelicolor* détectées par cette analyse, soient codés par la région centrale du chromosome de *S. ambofaciens*. Ces homologues forts seraient donc paralogues des produits codés par les régions terminales du chromosome de *S. ambofaciens*.

L'acquisition de séquences par transfert horizontal est également suggérée par l'analyse de séquences génomiques, notamment de leur pourcentage en bases G et C. En effet, les *Streptomyces* présentent un G-C% extrême avec une moyenne de 61%-79,7%. Ce biais de la composition de l'ADN se reflète dans la composition en bases des codons, avec en troisième position un taux de G et C variant de 76,4% à 98,3%. Sur cette base, chez *S. coelicolor* A3(2), l'analyse a révélé qu'un certain nombre de loci provenaient vraisemblablement d'événements de transfert horizontal. Ces loci correspondent à des éléments transposables, des gènes du métabolisme des hydrates de carbone, des méthylases d'ADN, des gènes de résistance aux métaux lourds ou des gènes codant des polycétone synthétases (Bentley *et al.*, 2002).

Cependant, il est probable que le nombre des gènes ayant subi un transfert soit fortement sous-estimé. En effet, parmi les genres bactériens occupant le même biotope que les *Streptomyces* se trouvent notamment *Rhodococcus* et *Mycobacterium*. Le génome de ces bactéries présente également un fort pourcentage en bases G et C avec respectivement 62%-70% et 62-66%. De plus, des plasmides linéaires, vecteurs potentiels du transfert horizontal, présentant la structure invertronicque de réplicons des *Streptomyces*, sont également retrouvés chez ces genres bactériens. Ainsi, il est plausible que des échanges entre espèces de caractéristiques génomiques proches aient lieu à fortes fréquences. Une fois transférées, les séquences présentant une valeur

adaptative, pourraient être exprimées après un processus d'amélioration rapide de la séquence pour atteindre le biais optimal de la composition en bases G et C.

### B-I.2.3.1.2 – Mécanismes de transfert d'information génétique aux extrémités des chromosomes des *Streptomyces*

Le mode de transfert d'information génétique provenant des régions subtélomériques est suggéré par l'observation d'interactions entre plasmides linéaires et chromosomes chez *S. griseus* et *S. coelicolor*.

Chez *S. griseus*, cette interaction conduit à un plasmide linéaire chimérique portant des séquences dérivant d'un bras chromosomique chez *S. rimosus* (Pandza *et al.*, 1998). Ce plasmide pPZG101 possède notamment, après recombinaison, les gènes de synthèse et de résistance à l'oxytétracycline précédemment portés par le chromosome en position subterminale.

Chez *S. coelicolor*, cette interaction aboutit à l'intégration de la partie centrale du plasmide linéaire SCP1 (Hanafusa and Kinashi, 1992).

Un autre indice de l'interaction entre plasmides linéaires et chromosomes chez les *Streptomyces* est donné indirectement par la comparaison des extrémités du chromosome de *S. lividans* et du plasmide linéaire SLP2. En effet, environ 16 kb d'ADN à l'extrémité du chromosome de *S. lividans* sont identiques à l'extrémité droite du plasmide linéaire SLP2 de 50 kb (Figure D-9) (Bey *et al.*, 2000).

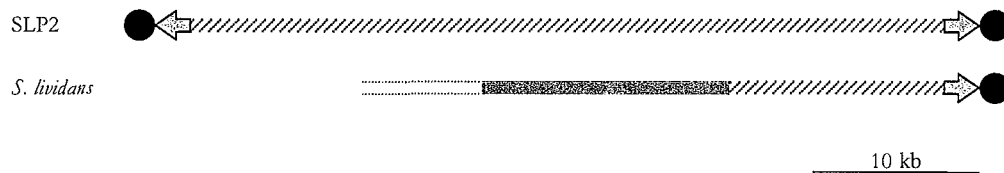


Figure D-9 : Comparaison du plasmide SLP2 et de l'extrémité droite du chromosome de *S. lividans*.

Les répétitions terminales inversées de SLP2 (flèches) ne sont pas représentées à l'échelle. Elles mesurent 44 pb et celles du chromosome de *S. lividans*, 30 kb. Les cercles noirs représentent les protéines terminales liées de façon covalente aux extrémités 5'. Les traits pointillés indiquent le début de la région subtélomérique droite du chromosome de *S. lividans*.

De plus, le clonage et séquençage des régions subterminales des chromosomes de *S. ambifaciens* ATCC23877 et DSM40697, non représentées dans les banques génomiques, a confirmé ces hypothèses (Gallois, 2003).

En effet, des ORF impliquées dans des fonctions de transfert conjugatif ont été identifiées aux extrémités du chromosome de la souche DSM40697 (Figure D-10). La première, *orf2*, coderait une protéine de 62 acides aminés présentant une identité significative avec les produits des gènes *spdA* des plasmides circulaires conjugatifs pJV1 de *S. phaeochromogenes* (Servin-Gonzalez *et al.*, 1995) et pSNA1 de *S. natalensis* ATCC27448 (Mendes *et al.*, 2000). Cette famille de protéines est impliquée dans le transfert intramycélien des plasmides conjugatifs. En effet, la conjugaison bactérienne chez *Streptomyces* peut être divisée en deux étapes, le transfert intermycélien contrôlé



### B-I.2.3.1.3 – Interactions plasmide/chromosome et acquisition de la linéarité chromosomique

Des approches de génomique comparée sur des espèces à chromosome linéaire et circulaire appartenant à un même phylum ont permis de mettre en évidence deux mécanismes hypothétiques d'acquisition de la linéarité des chromosomes chez les bactéries, notamment dans le cas d'*A. tumefaciens* (Wood *et al.*, 2001) et de *S. coelicolor* A3(2) (Bentley *et al.*, 2002).

#### B-I.2.3.1.2.1 - Interaction entre un plasmide linéaire et un chromosome circulaire ancestral chez *S. coelicolor* et *B. burgdorferi*

L'intégration d'un plasmide linéaire dans un chromosome circulaire constitue un mécanisme possible d'acquisition de la linéarité chez les Actinomycétales. La comparaison de la séquence de *S. coelicolor* A3(2) au chromosome linéaire avec celle de *Mycobacterium tuberculosis* (Cole *et al.*, 1998) dont le chromosome est circulaire appuie cette hypothèse (Bentley *et al.*, 2002).

Une conservation de l'ordre des gènes est observée entre la région « core » (environ 4,9 Mb) du chromosome de *S. coelicolor* et la totalité du génome de *M. tuberculosis* (4,4 Mb). Cette observation suggère que l'ancêtre commun des deux espèces possédait un chromosome circulaire d'environ 4 Mb. L'intégration d'un plasmide linéaire dans ce chromosome circulaire aurait conduit à l'apparition d'un chromosome linéaire, ancêtre du chromosome des *Streptomyces* (Figure D-11). Ainsi, un événement de recombinaison entre un plasmide linéaire et un chromosome circulaire ancestral conduit à l'apparition d'un chromosome linéaire dont les extrémités correspondent aux séquences du plasmide linéaire. Ainsi, les régions composant les « bras chromosomiques » de *S. coelicolor* auraient été acquises depuis la séparation évolutive entre les deux espèces, par intégration d'un plasmide linéaire puis divers réarrangements successifs résultant en la présence de 4 Mb d'ADN terminal hypervariable.

Un tel scénario prévoit que les fonctions de réplication des télomères soient situées à proximité des régions terminales du chromosome linéaire ancestral, dans les séquences spécifiques du plasmide (traits rouges). Ces fonctions de réplication, codées notamment par les gènes *tpg*, sont retrouvées à environ 100 kb de l'une des deux extrémités du chromosome de *S. coelicolor* A3(2) et de *S. lividans* (Bao and Cohen, 2001). De plus, des homologues non fonctionnels sont retrouvés à 8 kb des extrémités du chromosome de *S. lividans* (Yang *et al.*, 2002).

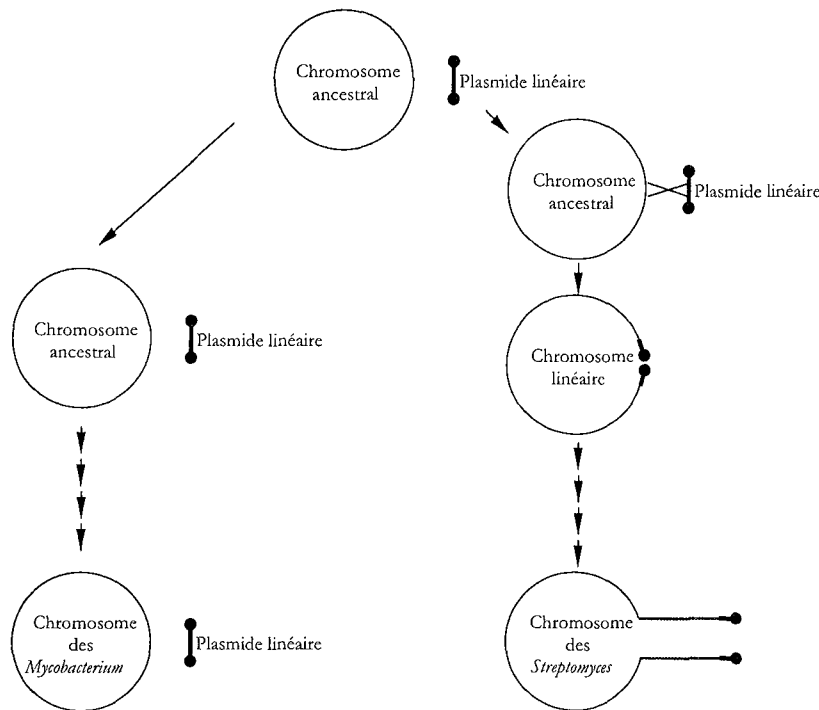


Figure D-11 : Modèle d'apparition de la linéarité des chromosomes des *Streptomyces*, à partir de l'ancêtre commun aux *Streptomyces* et aux *Mycobacterium*.

Chez *Borrelia*, le mode d'apparition de la linéarité pourrait être équivalent. Un événement d'intégration d'un réplicon linéaire pourrait expliquer la structure du chromosome actuel. Un des « donneurs » potentiels pourrait être l'iridopoxvirus, virus de la fièvre porcine africaine. En effet, l'iridopoxvirus et les *Borrelia* sont transmis par le même arthropode *Ornithodoros moubata*. Au sein de cet organisme, une intégration du génome viral dans le chromosome ancestral de *Borrelia* aurait pu avoir lieu, conduisant à la linéarisation du chromosome (Casjens, 1999).

#### B-I.2.3.1.2.2 - Acquisition de fonctions chromosomiques sur un plasmide linéaire ancestral chez *Agrobacterium tumefaciens*

*A. tumefaciens* possède deux chromosomes, l'un circulaire (3000 kb) et l'autre linéaire (2100 kb). Ce dernier serait un plasmide ayant acquis des séquences chromosomiques par transfert intragénomique. Trois arguments, issus de l'analyse de la séquence du chromosome linéaire, viennent étayer cette hypothèse (Goodner *et al.*, 2001; Wood *et al.*, 2001).

Ce chromosome linéaire possède une origine de répllication plasmidique similaire à celles de deux plasmides circulaires endogènes.

La proportion du nombre de gènes de "ménage" (impliqués dans la répllication et l'expression du génome) est plus importante sur le chromosome circulaire que sur le chromosome linéaire. Cette observation suggère que le transfert de fonctions essentielles à la survie de la bactérie s'est déroulé d'un chromosome circulaire ancestral vers un réplicon linéaire.

La syntonie entre *A. tumefaciens* et *Sinorhizobium meliloti*, une autre protéobactérie  $\alpha$  (Figure I-1), possédant un chromosome circulaire, montre une bonne conservation de l'ordre des gènes entre le chromosome circulaire d'*A. tumefaciens* et le chromosome de *S. meliloti*. Cependant, les homologues d'environ 300 gènes de *S. meliloti* sont retrouvés sur le chromosome linéaire d'*A. tumefaciens*. Ces 300 gènes sont dans le même ordre sur le chromosome linéaire et sur celui de *S. meliloti*, bien qu'ils soient plus épars sur le chromosome linéaire d'*A. tumefaciens*. Ceci suggère donc qu'un plasmide linéaire ancestral a acquis une partie des fonctions codées par le chromosome circulaire ancestral de *S. meliloti* et du chromosome circulaire de *A. tumefaciens*.

### B-I.2.3.2 – LES EXTREMITES TERMINALES INVERSEES DES CHROMOSOMES DE STREPTOMYCES ONT-ELLES UN ROLE ?

#### B-I.2.3.2.1 - Implication de l'instabilité dans l'évolution rapide des extrémités des *Streptomyces*

La formation d'hybrides chromosomiques ou plasmidiques par remplacement d'extrémités suite à un événement de transfert horizontal pourrait constituer une structure intermédiaire évoluant vers la reconstitution de TIR. En effet, tous les réplicons linéaires de *Streptomyces*, à l'exception de *S. avermitilis*, possèdent des TIR à leurs extrémités. Ces TIR peuvent se réduire à la taille des palindromes terminaux comme sur le plasmide SLP2 de *S. lividans* (Chen *et al.*, 1993) ou sur le chromosome de *S. avermitilis* (Omura *et al.*, 2001). Mais dans la plupart des cas décrits, elles couvrent plusieurs dizaines voire centaines de kilobases (voir paragraphe II.2.3.1.2. de l'introduction).

Ainsi, il est vraisemblable que la recombinaison terminale observée au niveau intraclonal engendre l'homogénéisation des régions terminales à partir d'un réplicon « hybride » (c'est-à-dire où les TIR originelles ont disparu suite à un événement d'échange d'extrémité). Cette homogénéisation ne serait cependant pas obligatoire, puisque le chromosome de *S. avermitilis* ne présente pas cette structure. Un échange d'extrémité ectopique, au niveau intragénomique, permettrait de recouvrir une structure invertronique, comme dans le cas des mutants spontanés NSAH135 et NSA855, issus de l'instabilité génétique de *S. ambifaciens*. En effet, ce type de réarrangement provoque la perte d'une extrémité et son remplacement par des séquences identiques à l'autre extrémité.

Ce modèle d'évolution rapide des régions terminales, par acquisition puis homogénéisation, est compatible avec la structure des TIR du chromosome de *S. lividans* 66 (Figure D-12).

Un premier événement de transposition, via l'élément transposable Tn4811, situé à 16 kb de l'extrémité du plasmide SLP2, aurait généré un chromosome hybride, possédant 16 kb de l'une des extrémités plasmidiques. Une cassure double brin, conduisant à la perte d'une des deux extrémités, conduit à l'apparition d'une extrémité recombinogène pouvant conduire à un



remplacement d'extrémité chromosomique et à l'homogénéisation des extrémités. La structure finale correspond à la structure des TIR de *S. lividans*, mesurant 30 kb, dont 16 kb homologues au plasmide SLP2 (Lin *et al.*, 1993; Redenbach *et al.*, 1993).

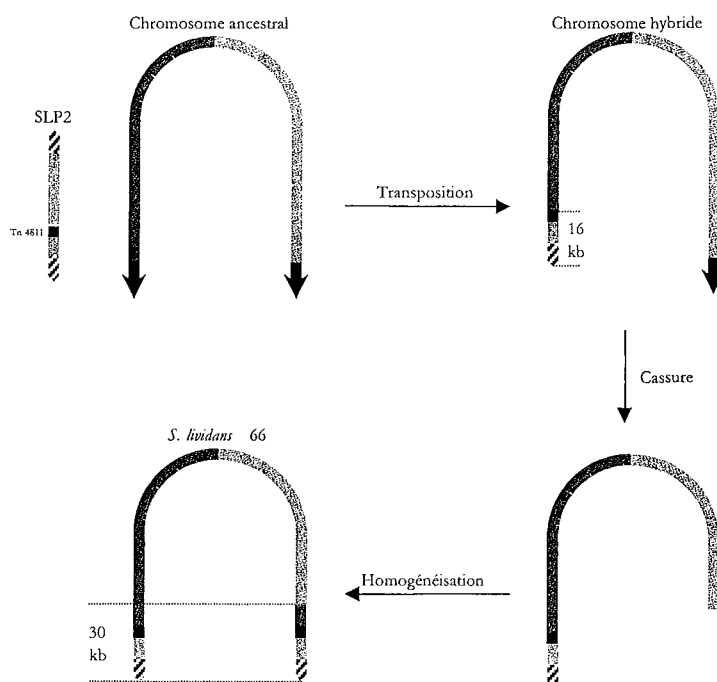


Figure D-12 : Hypothèse de formation des TIR du chromosome de *S. lividans* 66.

Les bras chromosomiques (droit en rouge, gauche en jaune) sont séparés par l'origine de réplication. Les répétitions terminales inversées (TIR pour Terminal Inverted Repeat) sont représentées par des flèches noires. Les protéines terminales ne sont pas représentées.

Ce modèle d'évolution rapide des extrémités permettrait également d'expliquer la grande diversité de la taille des TIR observée tant au niveau interspécifique (voir paragraphe II.2.3.1.2. de l'introduction) qu'au niveau intraclonal (résultats présentés dans ces travaux, Tableau D-1).

#### B-I.2.3.2.2 – Les TIR : un substrat pour la recombinaison terminale ?

La mise en évidence d'une grande variabilité de la taille des TIR lors de l'analyse de la descendance des souches *S. ambifaciens* NSA27 et NSA65 (Tableau D-1), a permis de révéler un autre aspect de l'instabilité génomique.

Certaines souches mutantes présentent des TIR de taille aussi réduite que 5 kb alors qu'elles mesurent 210 kb dans la souche sauvage. Cependant, dans les conditions de laboratoire, ces souches présentent un phénotype et un niveau d'instabilité de type sauvage.

Dérivées du sauvage	Taille des TIR (kb)	Dérivées de NSA27	Taille des TIR (kb)	Dérivées de NSA27	Taille des TIR (kb)	Dérivées de NSA65	Taille des TIR (kb)
NSAH135PP	480	NSA27 <sub>S18</sub>	370	NSA27 <sub>r7</sub>	5	NSA65 <sub>S19</sub>	105
NSA855	850	NSA27 <sub>S17S6'</sub>	480	NSA27 <sub>r10</sub>	80	NSA65 <sub>S19S3</sub>	1005
NSA27	6500	NSA27 <sub>S17S10'</sub>	480	NSA27 <sub>r11</sub>	85	NSA65 <sub>S4S9</sub>	1005
NSA65	6500	NSA27 <sub>S17S3'</sub>	850	NSA27 <sub>r4</sub>	85	NSA65 <sub>S4</sub>	6500
		NSA27 <sub>S17S9'</sub>	950	NSA27 <sub>r6</sub>	85	NSA65 <sub>S20</sub>	6500
		NSA27 <sub>S8</sub>	1140	NSA27 <sub>r1</sub>	110	NSA65 <sub>S4S7</sub>	6500
		NSA27 <sub>S13</sub>	1140	NSA27 <sub>r13</sub>	390		
		NSA27 <sub>S5</sub>	1300	NSA27 <sub>r14</sub>	390		
		NSA27 <sub>S20</sub>	1300	NSA27 <sub>r5</sub>	410		
		NSA27 <sub>S17S7'</sub>	1350	NSA27 <sub>r12</sub>	410		
		NSA27 <sub>S17S7'</sub>	1430	NSA27 <sub>r19</sub>	450		
		NSA27 <sub>S17S5'</sub>	1430	NSA27 <sub>r24</sub>	1400		
		NSA27 <sub>S17</sub>	6500				

Tableau D-1 : Taille des répétitions terminales inversées (TIR) des chromosomes des souches mutantes caractérisées au cours de ce travail.

L'information génétique codée par ces régions terminales ne semble donc pas être soumise à une forte pression de sélection. Ainsi, les répétitions terminales inversées du chromosome des *Streptomyces* pourraient ne jouer qu'un rôle structural, de substrat d'ADN nécessaire à la recombinaison terminale (voir paragraphe Perspectives).

- CHAPITRE B-II -

**DUPLICATIONS CHROMOSOMIQUES ET  
EVOLUTION**

## B-II.1 ☞ LA DYNAMIQUE DES GÉNOMES

### B-II.1.1 – LES *STREPTOMYCES* : MODELE D'ETUDE DE L'EVOLUTION DES GÉNOMES

Les bactéries du genre *Streptomyces* pourraient être considérées comme un modèle d'étude de l'évolution des génomes.

Les réarrangements caractérisés chez les souches mutantes issues de l'instabilité génétique de *S. ambofaciens*, pourraient en effet refléter les mécanismes d'évolution des *Streptomyces* dans leur écosystème naturel et ainsi révéler cette dynamique de l'évolution des génomes.

Les mutants NSA27 et NSA65 caractérisés au cours de ce travail présentent une fusion de chromosomes partiellement délétés. Ainsi, la taille de leur génome atteint les 13 Mb, en comparaison aux 8 Mb du chromosome de la souche de *S. ambofaciens*. Le contenu génique de ces souches est donc quasiment doublé par rapport à la souche sauvage.

Les souches NSAH135 et NSA855, au même titre que les descendants des souches NSA27 ou NSA65, présentent des remplacements de bras chromosomiques résultant d'évènements de recombinaison homologue entre les gènes dupliqués *has*. Le réarrangement présent chez ces souches mutantes consiste en la délétion de l'extrémité d'un bras chromosomique et en son remplacement par l'extrémité de l'autre bras chromosomique. Des régions de 480 kb ou 850 kb ont ainsi été dupliquées chez ces souches.

Dans la descendance de ces souches NSA27 et NSA65, une très grande variabilité chromosomique est observée, suggérant ainsi que les fusions chromosomiques, instables, génèrent divers réarrangements chromosomiques additionnels. Cette caractéristique permet de qualifier ces souches aux chromosomes fusionnés de mutatrices, générant une très grande variabilité génétique (voir paragraphe Perspectives).

De plus, l'importance du transfert horizontal dans l'évolution est également illustrée chez la bactérie *Streptomyces*, avec le polymorphisme des extrémités du chromosome qui la caractérise (voir paragraphe I.2.3 de la discussion).

Les mutants spontanés issus de l'instabilité génétique des *Streptomyces* pourraient donc refléter les mécanismes d'évolution des génomes, notamment via des grandes duplications de matériel génétique.

## B-II.1.2 - EVOLUTION DES GÉNOMES : TRANSFERT HORIZONTAL ET DUPLICATION GENIQUE

Comparativement aux eucaryotes, la taille des génomes bactériens présente une moindre variabilité. En effet, alors que les génomes eucaryotes peuvent varier en taille de quatre ordres de grandeur (entre environ  $10^7$  et  $10^{11}$  pb), la taille des génomes bactériens présente une variabilité de seulement un ordre de grandeur (entre environ  $10^6$  et  $10^7$  pb ; (Casjens, 1998)).

Cependant, cette différence au niveau de la gamme de tailles de génomes n'est pas à corrélérer avec le contenu en gènes. Par exemple, alors que les génomes de l'homme et de la levure ont une taille qui diffère de 300 fois, leur contenu en gènes n'est que 6 fois différent. A l'inverse, parmi les bactéries dont le génome est actuellement séquencé, les différences de taille de génomes reflètent des différences similaires dans leur nombre de gènes (de l'ordre de 10 fois). Ainsi, la variation de la taille d'un génome bactérien est directement liée à la complexité biochimique et physiologique de l'organisme, de par le fait que la majorité des séquences composant un génome bactérien est constituée de séquences codantes (Mira *et al.*, 2001).

La grande variabilité observée au niveau des propriétés métaboliques, des structures cellulaires ou encore de la différenciation des bactéries, en opposition avec une faible variabilité de la taille des génomes, peut alors être expliquée par une grande dynamique des génomes bactériens permettant l'évolution rapide de l'information génétique.

Selon Mira *et al.* (2001), les génomes bactériens évoluent grâce à un équilibre entre l'acquisition de nouvelles informations génétiques grâce à des événements de duplications ou de transfert horizontal et élimination de l'information ne conférant pas d'avantage adaptatif par formation de pseudogènes et de grandes délétions génomiques (Figure D-13).

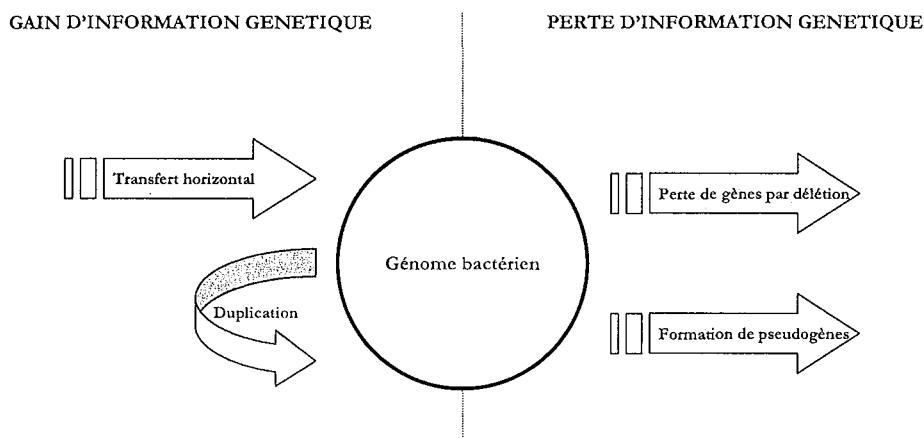


Figure D-13 : Modèle d'évolution de la taille des génomes bactériens (selon Mira *et al.*, 2001).

Selon Ochman *et al.* (2000), bien qu'il existe de nombreuses évidences de duplications géniques chez les génomes bactériens, la principale source d'acquisition d'information génétique chez les bactéries est le transfert horizontal. Selon les auteurs, les bactéries ont acquis une grande partie de

leur diversité génétique via l'acquisition de séquences par transfert horizontal, contrairement aux eucaryotes qui évoluent principalement à travers la modification de l'information génétique existante (même si le transfert horizontal a également été démontré chez divers eucaryotes). Le transfert horizontal serait à l'origine d'une grande dynamique des génomes, dans lesquels des quantités substantielles d'ADN seraient introduites et/ou délétées.

Par ailleurs, l'hypothèse de l'évolution par duplication de gènes, suggérée à l'origine par Ohno en 1970 (Ohno, 1970), est universellement acceptée. La duplication génique et l'apparition corrélée des familles de gènes joueraient un rôle important dans l'évolution moléculaire des génomes. Après duplication, les différentes copies d'un gène peuvent diverger et/ou acquérir des régulations différentielles qui peuvent par la suite aboutir à l'apparition de nouvelles fonctions. Des gènes paralogues ont en effet été observés dans tous les génomes séquencés jusqu'alors (Koonin *et al.*, 1997) et constituent une grande fraction des gènes identifiés (ca. 40% dans le cas de *Saccharomyces cerevisiae*, (Blandin *et al.*, 2000) ; 1/3 d'un génome eucaryote selon Wagner (2001)). Bien que des événements isolés de duplications géniques expliquent la présence de dupliquats, de plus en plus de données suggèrent que de nombreux gènes dupliqués résultent d'événements de duplication de grandes régions chromosomiques. Des duplications de la totalité du génome ont même été suggérées.

L'analyse de la séquence du génome de la levure *S. cerevisiae* a largement contribué à ce débat. En effet, des études menées dans la fin des années 1990 (Coissac *et al.*, 1997; Wolfe and Shields, 1997) ont montré que le génome de *S. cerevisiae* portait 55 segments chromosomiques, la plupart pouvant être regroupés en paires, contenant des séries de gènes paralogues dans le même ordre et la même orientation. Ces segments couvrent environ la moitié du génome de *S. cerevisiae* et représentent 400 paires de gènes paralogues. De tels segments chromosomiques pourraient être le reflet de duplications successives de grands blocs chromosomiques, de la duplication d'un chromosome entier voire même de la totalité du génome, tous ces événements étant suivis de délétions successives de gènes et de translocations réciproques. Quel que soit le mécanisme, environ 1/3 des gènes paralogues identifiés dans le génome de la levure sont retrouvés dans ces segments chromosomiques.

Selon l'hypothèse de la duplication du génome entier de Wolfe and Shields en 1997 (Wolfe and Shields, 1997), un événement d'allo- ou d'auto-tétraploïdisation se serait produit précisément après la divergence entre *Saccharomyces kluyveri* and *S. cerevisiae* (Figure D-12).

Une étude sur la redondance des gènes dans les génomes d'Hémiascomycètes (Llorente *et al.*, 2000a) a controversé cette hypothèse puisqu'elle a montré l'existence de la plupart des paires de gènes paralogues chez des espèces ayant émergé avant l'événement de duplication de génome proposé par Wolfe and Shields (Wolfe and Shields, 1997), notamment chez *Kluyveromyces lactis* (voir arbre phylogénétique, Figure D-14).

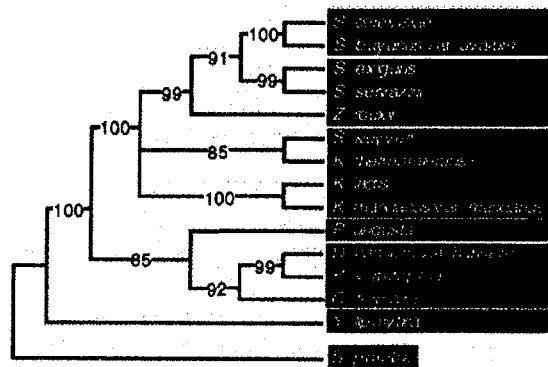


Figure D-14 : Arbre phylogénétique des 13 espèces d'Hémiascomycètes étudiés par Llorente *et al.* (2000).

Selon Souciet *et al.*, 2000. L'arbre est réalisé à partir des séquences codant les ARNr 23S par la méthode du maximum de parcimonie.

Cette étude a donc révélé la conservation de familles de gènes présentes dans le dernier ancêtre commun des 13 espèces d'Hémiascomycètes mais également l'apparition ou la disparition d'autres familles de gènes suite à la spéciation (Llorente *et al.*, 2000a).

Une analyse comparative de la syntenie entre *S. cerevisiae* et ces 13 espèces d'Hémiascomycètes a permis aux auteurs de proposer un nouveau modèle d'évolution des génomes de levure (Llorente *et al.*, 2000b). Selon ce modèle, le génome des levures aurait évolué par des événements de duplications de segments chromosomiques regroupant quelques gènes plutôt que par la duplication de chromosomes entiers ou d'autres réarrangements chromosomiques n'impliquant pas des duplications de séquences. Des délétions et translocations réciproques ultérieures permettraient d'expliquer la composition finale en gènes paralogues de chaque espèce (Llorente *et al.*, 2000b).

## B-II.2 $\propto$ LE DEVENIR DES GENES DUPLIQUES *HASR* ET *HASLAU* COURS DE L'EVOLUTION

L'étude des facteurs  $\sigma^{\text{Has}}$  et la caractérisation de leur rôle permettent d'aborder la thématique de l'évolution des gènes dupliqués. Deux hypothèses non exclusives pourraient expliquer la présence de la duplication des gènes *has* chez *Streptomyces ambofaciens* :

- La duplication des gènes *has* pourrait être récente et donc spécifique à *S. ambofaciens*.
- Alternativement, la duplication des gènes *has* pourrait être plus ancienne en terme d'évolution et leur forte conservation révélerait l'existence d'une pression de sélection et/ou d'un mécanisme d'homogénéisation de leurs séquences.

Les résultats obtenus lors de ce travail permettent d'une part d'estimer la « datation » à l'échelle évolutive de cette duplication, et d'autre part de suggérer l'existence d'une pression de sélection pour la conservation des duplicats par la mise en évidence d'une régulation différentielle des gènes *has*.

### B-II.2.1 – UN MECANISME D'EVOLUTION CONCERTEE A L'ORIGINE DE LA CONSERVATION DE LA DUPLICATION DES GENES *HAS* ?

L'analyse des séquences des génomes de deux espèces de *Streptomyces* (*S. coelicolor* et *S. avermitilis*) suggère fortement qu'un grand nombre de gènes serait apparu suite à des événements de duplications géniques durant l'évolution du genre *Streptomyces* (Ikeda et al., 2003). La famille des facteurs  $\sigma$  en particulier suggère cette évolution par duplication, puisque 65 et 60 facteurs  $\sigma$  ont été identifiés chez *S. coelicolor* et *S. avermitilis* respectivement, notamment en ce qui concerne la famille de 9 facteurs  $\sigma$  de type  $\sigma^{\text{B}}$ , qui est retrouvée aussi bien chez *S. coelicolor* que chez *S. avermitilis*.

Une analyse phylogénétique menée sur ces facteurs  $\sigma$  semble suggérer que la situation de *S. ambofaciens* est différente de celle de *S. coelicolor* et *S. avermitilis* puisque cette famille semble être composée de 10 membres chez *S. ambofaciens*, deux d'entre eux (les facteurs  $\sigma^{\text{Has}}$ ) présentant 98% d'identité nucléotidique.

Des expériences d'hybridation moléculaire, à l'aide d'une sonde correspondant à la partie codante des gènes *has*, révèlent en effet d'autres signaux que ceux correspondant aux gènes *has* eux-mêmes. De plus, des expériences de PCR réalisées à l'aide d'un couple d'oligonucléotides spécifiques du gène *sigL* de *S. coelicolor* (ne permettant pas d'amplifier les gènes *has*) ont révélé la présence d'un homologue de ce gène chez *S. ambofaciens*. Ces résultats suggèrent donc l'existence



de la famille entière des 9 gènes codant les facteurs  $\sigma$  de type  $\sigma^B$  sur le chromosome de *S. ambifaciens*.

La duplication des gènes *has* semblerait donc spécifique de l'espèce *S. ambifaciens* et apparaîtrait ainsi comme un événement récent en terme d'évolution.

Le séquençage des deux copies *has* chez les trois souches de *S. ambifaciens* (DSM40697, ETH12567, ATCC23877) a permis de montrer que la duplication des gènes *has* était présente au sein de l'espèce *S. ambifaciens*. En effet, les pourcentages d'identité nucléotidique entre les deux gènes *hasR* et *hasL* sont de 97% chez les souches ETH12567, ATCC23877, en comparaison aux 98% de la souche DSM40697.

Cette forte conservation de séquences n'est pas sans rappeler le cas des gènes dupliqués *tuf* chez les eubactéries, codant les facteurs d'élongation de la traduction EF-Tu. En effet, comme pour les gènes *hasR* et *hasL*, lorsque les deux gènes *tufA* et *tufB* sont présents sur un chromosome bactérien, leur localisation chromosomique est distante (Sela *et al.*, 1989). Lorsque la duplication est retrouvée chez une espèce bactérienne, les deux gènes *tuf* partagent une très forte identité nucléotidique (par exemple 99% chez *S. typhimurium* ; (Jaskunas *et al.*, 1975), et *E. coli* ; (Furano, 1978). Une analyse comparative des séquences des gènes *tuf* chez *S. typhimurium* et *E. coli* a suggéré que les gènes *tufA* et *tufB* chez chaque espèce n'évoluaient pas indépendamment l'un de l'autre (Sharp, 1991). Un mécanisme d'évolution concertée, tel que la conversion génique, permettrait d'aboutir à l'homogénéisation des séquences nucléotidiques des gènes *tuf*.

Les alignements des séquences des gènes *has* chez les trois isolats de *S. ambifaciens* ne révèlent cependant pas suffisamment de positions informatives pour pouvoir évoquer un tel phénomène d'évolution concertée dans le cas des gènes *has*.

## **B-II.2.2 – L'AVANTAGE ÉVOLUTIF DE LA DUPLICATION DES GÈNES *HAS***

L'impact de la duplication des gènes *has* sur l'information de l'espèce *S. ambifaciens* a été recherché au cours de ce travail par une analyse transcriptionnelle fine (RT-PCR quantitative) et la construction de mutants. En effet, il s'agissait d'une part d'associer un phénotype à la mutation de l'une voire des deux copies *has*. D'autre part, l'analyse transcriptionnelle était susceptible de révéler un contrôle transcriptionnel divergent. L'objectif à terme était de comprendre le rôle respectif des deux exemplaires de ce facteur  $\sigma$ , et donc d'envisager l'avantage sélectif, s'il existe, du maintien de cette duplication génique.

Préalablement à toute discussion concernant les raisons pouvant expliquer la conservation de la duplication des gènes *has* chez *S. ambifaciens*, il est nécessaire d'envisager la possibilité de l'évolution d'une des deux copies *has* vers la formation d'un pseudogène ou d'un gène codant une

protéine présentant une nouvelle fonction mais que la duplication soit trop récente à l'échelle évolutive pour visualiser cette dégénérescence.

Les résultats obtenus lors de ce travail permettent d'aborder les différents phénomènes pouvant expliquer l'existence d'une pression de sélection s'exerçant sur le maintien de la duplication des gènes *has*. Ainsi, la conservation des deux loci *has* au sein du génome de *S. ambofaciens* pourrait résulter de plusieurs phénomènes non exclusifs :

□ Les gènes *has* pourraient avoir divergé par modification de leur produit et acquis une fonction distincte. Ainsi, chaque facteur  $\sigma$  pourrait contrôler l'expression de régulateurs spécifiques. Ce mode de divergence s'apparente à la néofonctionnalisation.

Cependant les résultats obtenus lors de ce travail semblent montrer que les deux facteurs  $\sigma^{\text{Has}}$  contrôlent un même régulateur, impliqué dans la réponse au stress. Aucune différence significative n'a par ailleurs pu être mise en évidence lors de la recherche de phénotype chez les simples mutants.

Il est cependant intéressant de noter que la plupart des mutations nucléotidiques observées chez les deux autres souches de *S. ambofaciens*, comparativement à la souche DSM40697, sont silencieuses ou conservatives, c'est-à-dire qu'elle ne change pas la séquence protéique des facteurs  $\sigma^{\text{Has}}$  ou le change en un acide aminé similaire. Seuls deux changements sont observés en position 270 et 281 de la protéine HasR chez la souche *S. ambofaciens* ETH12567, dont un est conservatif. Enfin, les 8 différences observées entre les protéines HasR et HasL chez la souche *S. ambofaciens* DSM40697 sont retrouvées chez les deux autres souches (à l'exception du changement en position 281 chez la souche ETH12567), suggérant l'existence d'une pression de sélection pour la conservation de ces acides aminés spécifiques, qui pourraient donc révéler une réelle différence fonctionnelle entre les deux facteurs  $\sigma^{\text{Has}}$ .

□ La deuxième hypothèse permettant d'expliquer la conservation de la duplication des gènes *has* durant une longue période chez *S. ambofaciens* repose sur le modèle de Force (Force *et al.*, 1999). Les gènes *has* présenteraient une redondance fonctionnelle mais un environnement génétique et/ou des régions régulatrices spécifiques. Ils seraient donc exprimés de façon différentielle. Ce mode de divergence s'apparente à la sous-fonctionnalisation.

La divergence de l'environnement génétique des deux gènes *has*, avec notamment un système anti-anti- $\sigma$  / anti- $\sigma$  spécifique du gène *hasR*, semble en effet être en accord avec cette hypothèse. De plus, les résultats de RT-PCR quantitative révèlent un niveau d'expression très différent entre chaque copie *has*, puisque le gène *hasR* peut être jusqu'à 100 fois plus transcrit que le gène *hasL* chez la souche sauvage de *S. ambofaciens*. La poursuite des expériences visant à détecter l'activité des promoteurs *has in vitro* par détection de l'EGFP (voir paragraphe B-III.3.2 de la seconde partie des Résultats) permettra de conclure quant à la possibilité d'une expression différentielle des gènes *has* dans les différents compartiments cellulaires d'une colonie de *Streptomyces* (mycélium végétatif, mycélium aérien et spores).

Il est également important d'envisager qu'une différence d'expression entre les deux gènes *has* pourrait exister au niveau post-transcriptionnel voire même post-translationnel avec notamment

les système anti-anti- $\sigma$  / anti- $\sigma$  où les anti-facteur  $\sigma$ , codés par les gènes *prsR* et *prsL*, pourraient ne pas avoir la même affinité pour chacun des deux facteurs  $\sigma^{\text{HasR}}$  et  $\sigma^{\text{HasL}}$  (voir paragraphe A-I.2.3) ou encore la dégradation préférentielle de la protéine HasL médiée par les protéases de type Clp (voir paragraphe A-I.2.2).

□ Enfin, la perte d'une des deux copies *has* pourrait être contre-sélectionnée suite à l'évolution de chaque gène *has* vers une sous-fonction et l'accumulation de mutations dégénératives dans leurs régions promotrices (selon le modèle DCD de Force *et al.*, 1999). L'action combinée des deux loci *has* serait devenue nécessaire à l'accomplissement de la fonction ancestrale

Cette hypothèse prend d'autant plus de vraisemblance en terme de niveau d'expression relatif de chacun des deux gènes *has*. Nous pouvons effectivement imaginer que, bien que peu transcrit, le gène *hasL* soit nécessaire à la fonction codée par le gène *hasR*. Par exemple, un dosage fin des quantités de protéines Has par rapport aux quantités d'anti-facteur  $\sigma$  pourrait être à l'origine de la conservation de la duplication. Sans les protéines HasL, le facteur  $\sigma^{\text{HasR}}$  pourrait être en quantité insuffisante par rapport à la quantité d'anti facteur  $\sigma$  et ainsi ne pas pouvoir exercer sa fonction.

Cette hypothèse est en accord avec les résultats obtenus, puisque dans un mutant interrompu pour le gène *hasL*, le niveau de transcription du gène *hasR* chute considérablement. Cette diminution de la transcription reflète d'une part le contrôle croisé du gène *hasR* par le facteur  $\sigma^{\text{HasL}}$ . D'autre part, compte tenu du faible niveau de transcription du gène *hasL* et de l'auto-activation élevée du gène *hasR* par lui-même, cette forte diminution de la transcription du gène *hasR* dans un mutant  $\Delta\textit{hasL}$  ne peut être attribuée qu'à un niveau de régulation plus complexe et notamment à la titration du facteur  $\sigma^{\text{HasR}}$  par l'anti-facteur  $\sigma$ , ce dernier étant plus disponible de part l'absence de facteur  $\sigma^{\text{HasL}}$ .

La perte de la duplication des gènes *has* entraînerait donc une moindre activité du facteur  $\sigma^{\text{Has}}$  et serait donc contre-sélectionnée. Selon cette hypothèse, l'interruption de l'un comme l'autre des gènes *hasR* ou *hasL* entraînerait le même phénotype, dû à la perte de la fonction Has, ce qui est effectivement observé dans le cas des simples mutants. Cependant, l'interruption des deux gènes *has* devrait entraîner un phénotype, dû à la perte de la fonction Has. Or, les doubles mutants  $\Delta\textit{hasR}\Delta\textit{hasL}$  ne présentent pas de sensibilité accrue au stress oxydatif, comparativement aux simples mutants. L'absence de phénotype marqué ainsi que l'absence de différence de phénotype conféré par l'interruption de l'un ou l'autre ou encore des deux gènes *has*, peuvent être expliquées par la redondance des facteurs  $\sigma$  de type  $\sigma^{\text{B}}$ . Ainsi, la fonction ancestrale serait pleinement assurée par l'ensemble des facteurs  $\sigma$  composant cette famille.

C - TROISIEME PARTIE :

PERSPECTIVES

## C-I ☞ VALEURS ADAPTATIVES DES DUPLICATIONS TERMINALES

L'analyse de la descendance des souches mutantes issues de l'instabilité génétique de *S. ambofaciens* NSA27 et NSA65 a permis de révéler l'existence d'un phénomène générateur de diversité génétique. Ainsi, les souches portant un chromosome fusionné constituent des « mutateurs chromosomiques », à l'origine d'une grande variabilité de structures chromosomiques, avec notamment des duplications de grandes régions génomiques. Cet état mutateur, hautement instable, pourrait être transitoire, dans le sens où il permettrait l'apparition de structures génomiques variées pouvant présenter un avantage adaptatif au regard de l'évolution.

Mais quelle serait la nature de cet avantage, s'il existe, conféré par des duplications de séquences chromosomiques terminales ?

Deux hypothèses non exclusives peuvent être énoncées.

(i) L'avantage adaptatif des duplications terminales peut résulter de la duplication de séquences codantes. L'évolution des deux copies pouvant être à l'origine de nouvelles fonctions. L'analyse de l'expression des gènes dupliqués pourrait être étendue aux gènes présents dans les TIR. L'accès à la technologie des filtres à haute densité (« macroarrays ») réalisés à partir des séquences des TIR du chromosome de *S. ambofaciens* ATCC23877 (obtenus en collaboration avec le Génoscope et l'IGM d'Orsay), permettra à court terme d'aborder ce problème de façon beaucoup plus systématique.

(ii) L'avantage pourrait également résulter de la présence d'une structure répétée dans les régions terminales plus que de la capacité codante. En effet, les répétitions terminales inversées (TIR) et autres séquences répétées présentes dans les régions subtélomériques, peuvent apparaître comme une conséquence de la recombinaison mais également comme des substrats permettant la recombinaison terminale et donc la réparation de cassures chromosomiques. Ces structures participeraient ainsi au maintien de l'intégrité chromosomique.

L'avantage de la présence de répétitions terminales et subtélomériques pourrait être testé par l'analyse de l'instabilité phénotypique et moléculaire des différents mutants présentant des duplications terminales de tailles variables. La présence de TIR de taille variée pourrait être corrélée à la capacité de résister aux agents cassants l'ADN, inducteurs de cassures double brin.

La duplication des gènes *has*, localisés dans les régions terminales du chromosome, constitue un modèle expérimental pour étudier l'impact des duplications génomiques. Cette duplication est récente puisqu'absente des espèces *S. avermitilis* et *S. coelicolor*, cette dernière étant étroitement apparentée à *S. ambofaciens*. De plus, elle accompagne un réarrangement terminal de grande ampleur, translocation d'au moins 160 kb d'un bras chromosomique vers l'autre. Bien que la duplication soit récente, l'analyse des séquences des gènes *has* chez les trois isolats de

*S. ambofaciens* suggère que la fonction des deux copies *has* est soumise à pression de sélection. De plus, leur analyse fonctionnelle suggère fortement que les deux copies sont fonctionnelles et participeraient à une régulation fine de la réponse générale au stress.

Quel pourrait être l'avantage de la duplication d'un gène codant un facteur sigma ?

L'analyse fine de la réponse au stress de souches et d'espèces ne présentant qu'une seule copie *has* pourrait apporter des éléments dans ce sens. L'interruption du gène *sigB* chez *S. coelicolor*, homologue des gènes *has*, confère notamment un phénotype de sensibilité au stress osmotique. L'absence de cette sensibilité chez les mutants interrompus pour le(s) gène(s) *has* révèle-t-elle l'avantage de la duplication des gènes *has* ? La souche *S. ambofaciens* aurait-elle acquis une résistance accrue face au stress osmotique, notamment par rapport à *S. coelicolor*, suite à la duplication du gène codant le facteur  $\sigma$  de réponse à ce stress ? Si tel était le cas, une souche de *S. ambofaciens* ne codant plus que pour une seule copie de ce facteur  $\sigma$  devrait être désavantagée par rapport à la souche sauvage. Ainsi, des expériences de compétitions entre la souche sauvage de *S. ambofaciens* et un mutant interrompu pour l'un ou l'autre des gènes *hasR* ou *hasL*, devraient permettre de distinguer ces deux souches en conditions de stress osmotique notamment.

## C-II $\&$ CONTROLE DE LA REPONSE AU STRESS CHEZ LES *STREPTOMYCES*

Outre leur présence en deux copies, les gènes *has* présentent l'intérêt de coder des régulateurs de la transcription. Ainsi, les résultats obtenus permettent de mieux comprendre les mécanismes de réponse aux stress chez les *Streptomyces*.

### C-II.1 – CONTROLE DE L'EXPRESSION DES GENES *HAS*

#### C-II.1.1 – CONTROLE LA TRANSCRIPTION DES GENES *HAS*

L'analyse des contextes mutants *has* a permis de montrer que différents facteurs sigma (dont les facteurs  $\sigma^{\text{Has}}$ ) pouvaient vraisemblablement rentrer en compétition. Les phénomènes d'inhibition pourraient traduire ces interactions compétitives de natures différentes.

L'inhibition peut tout d'abord résulter d'un phénomène d'encombrement stérique au site promoteur. Ainsi, la région promotrice *hasR* comprend deux sites de fixation pour le facteur  $\sigma$  WhiG chevauchant le promoteur *sigB*-like. La fixation de WhiG pourrait diminuer l'accès du promoteur  $\sigma^{\text{Has}}$  dépendant. Afin de tester cette hypothèse, l'introduction en multi-copies de la région promotrice du gène *hasR* pourrait être réalisée. En effet, si ce promoteur putatif était effectivement reconnu par le facteur  $\sigma$  WhiG, son introduction en multi-copies entraînerait la titration de ce facteur  $\sigma$  et donc l'apparition d'un phénotype de différenciation (Tan and Chater, 1993).

Chez *S. coelicolor* A3(2), une relation épistatique a été établie par l'étude de l'expression de *sigF*, impliqué dans la maturation des spores, chez un mutant *whiG* (Kelemen *et al.*, 1996). L'expression de *sigF* est dépendante de la présence de  $\sigma$  WhiG actif. Cependant, aucune séquence spécifique de reconnaissance par  $\sigma$  WhiG en amont du gène *sigF* n'a été mise en évidence. Ainsi, un facteur  $\sigma$  est manquant dans la cascade de contrôle menant à la différenciation, entre les facteurs  $\sigma$  WhiG et  $\sigma^{\text{F}}$ . Ainsi, le facteur  $\sigma^{\text{HasR}}$  et son homologue chez *S. coelicolor* sont des candidats pour jouer ce rôle.

Dans ce cas, les facteurs  $\sigma^{\text{Has}}$  constitueraient un nouvel exemple de relation entre la différenciation morphologique et la réponse aux stress chez les *Streptomyces*.

Une seconde hypothèse consiste à impliquer l'ensemble des facteurs sigma  $\sigma^{\text{B}}$ -like (9 chez *S. coelicolor* et *S. avermitilis*, vraisemblablement 10 chez *S. ambofaciens*) dans le contrôle de tout ou partie des gènes du régulon, gènes codant les facteurs  $\sigma^{\text{Has}}$  compris. Ce réseau de régulation complexe pourrait intégrer l'interaction croisée des facteurs sigma avec les différents promoteurs du régulon. Ces interactions pourraient être modulées par l'affinité variable d'un facteur sigma pour des séquences promotrices de séquences flexibles. De plus, la disponibilité de chaque facteur sigma pourrait également résulter de l'interaction avec les différents anti-facteurs  $\sigma$  associés. Afin d'apporter des éléments de réponse, l'ensemble des gènes codant les facteurs

sigma de la famille  $\sigma^B$ -like doit être étudié par interruption au locus chromosomique ou surexpression depuis un vecteur d'expression (voir paragraphe C-II.2).

### C-II.1.2 – LES ANTI-FACTEURS $\sigma$ CONTROLANT L'ACTIVITE DES FACTEURS $\sigma^{HAS}$

La régulation post-traductionnelle des gènes *has* reste à étudier. Un mutant interrompu pour le gène *prsR* a été construit par la méthode de « PCR targeting » (Gust *et al.*, 2003) et l'analyse par PCR quantitative du niveau d'expression de chaque transcrite *has* dans ce contexte est actuellement en cours. En effet, l'activité de chaque facteur  $\sigma^{HAS}$  pourra être modifiée dans un contexte mutant pour le(s) gène(s) codant l'anti- $\sigma$ . Le niveau transcriptionnel des gènes *has* sera altéré par voie de conséquence. L'interruption des gènes *prsL*, *prsR* et *arsR* permettra donc non seulement d'aborder la complexité des systèmes de régulation de l'expression génique chez les *Streptomyces* mais également de détecter une éventuelle régulation différentielle des facteurs  $\sigma^{HAS}$  au niveau post-traductionnel.

## **C-II.2 – LE ROLE DU REGULON $\sigma^{HAS}$**

### C-II.2.1 – RECHERCHE DE PHENOTYPE CHEZ DES MUTANTS *HAS*

Les résultats obtenus ainsi que ceux de la littérature indiquent qu'une famille de gènes codant des facteurs sigma soit responsable de la réponse à un stress. En l'occurrence, la réponse générale aux stress pourrait être assurée par les 10 facteurs  $\sigma$  homologues chez *S. ambofaciens*. L'interruption successive des gènes codant cette famille de protéines et l'étude du phénotype de chaque mutant permettra de comprendre la part respective de chaque facteur  $\sigma$  dans la réalisation d'une réponse adaptée au stress. Cette analyse permettra peut-être de comprendre pourquoi la présence de 9 facteurs  $\sigma$  (voire 10 chez *S. ambofaciens*) a été sélectionnée chez les *Streptomyces*, alors que le seul facteur  $\sigma^B$  suffit à remplir un rôle équivalent chez *B. subtilis*.

Chez *S. coelicolor*, les mutants interrompus pour le gène *rsrA*, codant l'anti-facteur  $\sigma$  spécifique du facteur  $\sigma^R$ , sont caractérisés par une sporulation déficiente (Paget *et al.*, 2001a). Ce phénotype de différenciation, identique à celui observé chez les mutants du gène *whiG* (Chater *et al.*, 1989), pourrait refléter une compétition entre le facteur  $\sigma^R$  et le facteur  $\sigma$  WhiG pour le « core » enzyme de l'ARN polymérase. En effet, dans un contexte mutant *rsrA*, la quantité de facteur  $\sigma^R$  libre serait plus élevée que dans un contexte sauvage, entraînant donc une dérégulation du dosage des différentes sous-unités de l'ARN polymérase. La caractérisation du phénotype des mutants des anti-facteurs  $\sigma$  PrsR et PrsL pourrait apporter le même type d'information.

Des doubles mutants interrompus à la fois pour le gène *prsR* et également pour le gène *hasR* ou *hasL* ont également été construits. En effet, dans le cas du facteur  $\sigma^H$  mais aussi du facteur  $\sigma$  codé par le gène *sigU*, l'interruption simultanée des gènes codant l'anti-facteur  $\sigma$ ,



respectivement PrsH et RsuA, et le facteur  $\sigma$  associé entraîne l'apparition d'un phénotype de différenciation, à la différence du mutant du facteur  $\sigma$  seul (Gehring *et al.*, 2001; Viollier *et al.*, 2003b). De façon similaire, les anti-facteur  $\sigma$  PrsR et PrsL pourraient également contrôler l'activité d'autres facteurs  $\sigma$  que les facteurs  $\sigma^{\text{Has}}$  et leur interruption pourrait permettre d'accéder à cette complexité des systèmes de régulation des facteurs  $\sigma$  chez les *Streptomyces*.

### C-II.2.2 – IDENTIFICATION DES GENES CIBLES DES FACTEURS $\sigma^{\text{Has}}$

L'implication des facteurs  $\sigma^{\text{Has}}$  dans la réponse aux stress peut également être confirmée par l'identification de leurs gènes cibles.

La recherche de consensus de promoteurs chez les *Streptomyces* n'a permis de caractériser qu'environ 20% des gènes étudiés jusqu'alors (Bourn and Babb, 1995). Ainsi, il est probable que le nombre de gènes étudiés soit encore insuffisant pour définir des consensus et donc d'attribuer un mode de régulation transcriptionnelle à une ORF sur la seule base de l'analyse de la séquence de la région 5' promotrice. C'est en effet lors de la caractérisation de chaque facteur  $\sigma$  que le consensus des promoteurs reconnus par ce dernier peut être défini. Ainsi, chez *S. coelicolor*, des consensus de promoteurs reconnus par les facteurs  $\sigma^{\text{HrdB}}$ ,  $\sigma^{\text{WhiG}}$ ,  $\sigma^{\text{E}}$ ,  $\sigma^{\text{R}}$  et SigB ont été définis. Cependant, par rapport au grand nombre de facteurs  $\sigma$  codés dans le génome de cette espèce, il semble clair que les différents types de promoteurs chez *S. coelicolor* restent à définir. Une analyse bioinformatique visant à caractériser des combinaisons de séquences retrouvées dans les régions intergéniques de la séquence du génome de *S. coelicolor* A3(2) permettra de définir le niveau de conservation des promoteurs des gènes d'un même régulon (en cours de réalisation au laboratoire). Cette analyse a pour objectif à moyen terme de définir des réseaux de gènes sur la base de l'analyse de leur région promotrice.

En ce qui concerne les facteurs  $\sigma^{\text{Has}}$ , les séquences des promoteurs *hasRp1* et *hasLp1* sont très similaires aux promoteurs reconnus par le facteur de réponse générale au stress  $\sigma^{\text{B}}$  de *B. subtilis* et par le facteur  $\sigma$  SigB de *S. coelicolor* impliqué dans la réponse au stress osmotique (Cho *et al.*, 2001). La recherche des séquences similaires au consensus des promoteurs de type  $\sigma^{\text{Has}}$  dans les régions intergéniques des séquences des régions terminales du chromosome de *S. ambofaciens* permettra d'identifier les gènes cibles des facteurs  $\sigma^{\text{Has}}$  localisés dans ces régions. Dans l'attente de la séquence génomique de *S. ambofaciens*, les séquences similaires au consensus ANGNNNT-N<sub>14-16</sub>-GGGTA (défini par comparaison avec *S. coelicolor* et *B. subtilis*) ont été recherchées chez *S. coelicolor*. Ainsi, 20 gènes candidats ont pu être identifiés et leur transcription chez les mutants interrompus pour le(s) gène(s) *has* sont en cours d'analyse. Il est cependant intéressant de noter la présence, parmi ces gènes, de *catB*, codant une catalase. Ce gène est régulé par le facteur  $\sigma$  SigB chez *S. coelicolor* (Cho *et al.*, 2001) qui possède deux catalases monofonctionnelles : CatA et CatB. La catalase codée par le gène *catB*, sous contrôle du facteur  $\sigma$  SigB, a été impliquée dans la défense contre le stress osmotique ainsi que dans la différenciation de *S. coelicolor* (Cho *et al.*, 2000). La catalase CatA est quant à elle induite en conditions de stress oxydatif par le peroxyde d'hydrogène (Cho and Roe, 1997). Ainsi, la prédiction *in silico* de l'appartenance d'une catalase au

régulon  $\sigma^{\text{Has}}$  serait étayée par les résultats fonctionnels des gènes *has*, en particulier l'induction transcriptionnelle des gènes *has* dans des conditions de stress oxydatif et osmotique.

L'analyse de la transcription des gènes cibles candidats sera également réalisée dans les différents contextes mutants pour les 10 facteurs  $\sigma$  de la famille caractérisée, et permettra de déterminer les spécificités et le rôle respectif de chacun de ces facteurs  $\sigma$  dans la réponse générale aux stress.

## 2 BIBLIOGRAPHIE 6

- 1 **Aigle, B., D. Schneider, C. Morilhat, D. Vandewiele, A. Dary, A. C. Holl, J. M. Simonet and B. Decaris** (1996). An amplifiable and deletable locus of *Streptomyces ambofaciens* RP181110 contains a very large gene homologous to polyketide synthase genes. *Microbiology* **142**: 2815-2824.
- 2 **Allendorf, F. W., F. M. Utter and B. P. May** (1975). Gene duplication within the family *Salmonidae*: Detection and determination of the genetic control of duplicate loci through inheritance studies and the examination of populations. *Isozymes, Genetics and evolution*. New-York, Academic Press: 415-432.
- 3 **Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman** (1990). Basic Local Alignment Search Tool. *J Mol Biol* **215**(3): 403-10.
- 4 **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17): 3389-402.
- 5 **Altuvia, S., M. Almiron, G. Huisman, R. Kolter and G. Storz** (1994). The dps promoter is activated by OxyR during growth and by IHF and sigma S in stationary phase. *Mol Microbiol* **13**(2): 265-72.
- 6 **Angerer, A., S. Enz, M. Ochs and V. Braun** (1995). Transcriptional regulation of ferric citrate transport in Escherichia coli K-12. Fecl belongs to a new subfamily of sigma 70-type factors that respond to extracytoplasmic stimuli. *Mol Microbiol* **18**(1): 163-74.
- 7 **Ansari, A. and M. R. Gartenberg** (1999). Persistence of an alternate chromatin structure at silenced loci *in vitro*. *Proceedings of the National Academy of Sciences USA* **96**: 343-348.
- 8 **Aubert, M., E. Weber, B. Gintz, K. F. Chater and B. Decaris** (1997). Inactivation or amplification of the *spa2* gene, encoding a potential stationary-phase regulator, affects differentiation in *Streptomyces ambofaciens*. *Canadian Journal of Microbiology* **43**: 1118-1125.
- 9 **Aubert, M., E. Weber, D. Schneider, J. M. Simonet and B. Decaris** (1993). Primary structure analysis of a duplicated region in the amplifiable AUD6 locus of *Streptomyces ambofaciens* DSM40697. *FEMS Microbiology Letters* **113**: 49-56.
- 10 **Bancroft, I.** (2001). Duplicate and diverge: the evolution of plant genome microstructure. *Trends Genet* **17**(2): 89-93.
- 11 **Bao, K. and S. N. Cohen** (2001). Terminal proteins essential for the replication of linear plasmids and chromosomes in Streptomyces. *Genes Dev* **15**(12): 1518-27.
- 12 **Barne, K. A., J. A. Bown, S. J. Busby and S. D. Minchin** (1997). Region 2.5 of the Escherichia coli RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *Embo J* **16**(13): 4034-40.
- 13 **Bashyam, M. D. and A. K. Tyagi** (1998). Identification and analysis of "extended -10" promoters from mycobacteria. *J Bacteriol* **180**(9): 2568-73.
- 14 **Baudin, A., O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute and C. Cullin** (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **21**(14): 3329-30.
- 15 **Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D.**

extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* **35**(3): 490-516.

44 **Chater, K. F.** (1993). Genetics of differentiation in *Streptomyces*. *Annual Review of Microbiology* **47**: 685-713.

45 **Chater, K. F.** (1998). Taking a genetic scalpel to the *Streptomyces* colony. *Microbiology* **144**: 1465-1478.

46 **Chater, K. F., C. J. Bruton, K. A. Plaskitt, M. J. Buttner, C. Mendez and J. D. Helmann** (1989). The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility sigma factor of *B. subtilis*. *Cell* **59**(1): 133-43.

47 **Chen, C. W., T. W. Yu, Y.-S. Lin, H. M. Kieser and D. A. Hopwood** (1993). The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50 kb linear molecule. *Molecular Microbiology* **7**: 925-932.

48 **Cho, Y. H., E. J. Lee, B. E. Ahn and J. H. Roe** (2001). SigB, an RNA polymerase sigma factor required for osmoprotection and proper differentiation of *Streptomyces coelicolor*. *Mol Microbiol* **42**(1): 205-14.

49 **Cho, Y. H., E. J. Lee and J. H. Roe** (2000). A developmentally regulated catalase required for proper differentiation and osmoprotection of *Streptomyces coelicolor*. *Mol Microbiol* **35**(1): 150-60.

50 **Cho, Y. H. and J. H. Roe** (1997). Isolation and expression of the *catA* gene encoding the major vegetative catalase in *Streptomyces coelicolor* Muller. *J Bacteriol* **179**(12): 4049-52.

51 **Chu, G., D. Vollrath and R. W. Davis** (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**: 1582-1585.

52 **Chung, H. J., E. J. Kim, B. Suh, J. H. Choi and J. H. Roe** (1999). Duplicate genes for Fe-containing superoxide dismutase in *Streptomyces coelicolor* A3(2). *Gene* **231**(1-2): 87-93.

53 **Ciampi, M. S., M. B. Schmid and J. R. Roth** (1982). Transposon Tn10 provides a promoter for transcription of adjacent sequences. *Proceedings of the National Academy of Sciences USA* **79**: 5016-5020.

54 **Citti, C. and K. S. Wise** (1995). *Mycoplasma hyorhinis vlp* gene transcription: critical role in phase variation and expression of surface lipoproteins. *Molecular Microbiology* **18**: 649-660.

55 **Clark, R. L. and F. C. Neidhardt** (1990). Roles of the two lysyl-tRNA synthetases of *Escherichia coli*: analysis of nucleotide sequences and mutant behavior. *J Bacteriol* **172**(6): 3237-43.

56 **Coissac, E., E. Maillier and P. Netter** (1997). A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres. *Mol Biol Evol* **14**(10): 1062-74.

57 **Cole, S. and I. Saint-Girons** (1999). Bacterial genomes - All shapes and sizes. *Organization of the Prokaryotic Genome*. R. L. Charlebois. Washington, D.C., American Society for Microbiology: 35-62.

58 **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. Mclean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell**

(1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**(6685): 537-44.

59 **Connolly, L., A. De Las Penas, B. M. Alba and C. A. Gross** (1997). The response to extracytoplasmic stress in *Escherichia coli* is controlled by partially overlapping pathways. *Genes Dev* **11**(15): 2012-21.

60 **Cooke, J., M. A. Nowak, M. Boerlijst and J. Maynard-Smith** (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet* **13**(9): 360-4.

61 **Cowing, D. W., J. C. Bardwell, E. A. Craig, C. Woolford, R. W. Hendrix and C. A. Gross** (1985). Consensus sequence for *Escherichia coli* heat shock gene promoters. *Proc Natl Acad Sci U S A* **82**(9): 2679-83.

62 **Cronn, R. C., R. L. Small and J. F. Wendel** (1999). Duplicated genes evolve independently after polyploid formation in cotton. *Proc Natl Acad Sci U S A* **96**(25): 14406-11.

63 **Daniels, D., J. L. Schroeder, J. L. Blattner, W. Szybalski and F. Sanger** (1983). A molecular map of coliphage lambda. *Lambda II, Appendix I*. R. W. Hendrix, J. W. Roberts, F. W. Stahl and R. A. Weisberg. New-York, Cold Spring Harbor: 469-517.

64 **Dary, A., N. Bourget, N. Girard, J.-M. Simonet and B. Decaris** (1992). Amplification of a particular DNA sequence in *Streptomyces ambofaciens* RP181110 reversibly prevents spiramycin production. *Research in Microbiology* **143**: 99-112.

65 **Dary, A., P. Kaiser, N. Bourget, C. J. Thompson, J.-M. Simonet and B. Decaris** (1993). Large genomic rearrangements of the unstable region of *Streptomyces ambofaciens* are associated with major changes in global gene expression. *Molecular Microbiology* **10**: 759-769.

66 **Datsenko, K. A. and B. L. Wanner** (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**(12): 6640-5.

67 **Davis, N. K. and K. F. Chater** (1992). The *Streptomyces coelicolor whiB* gene encodes a small transcription factor-like protein dispensable for growth but essential for sporulation. *Molecular and General Genetics* **232**: 351-358.

68 **de Crécy-Lagard, V., P. Servant-Moisson, J. Viala, C. Grandvalet and P. Mazodier** (1999). Alteration of the synthesis of the Clp ATP-dependent protease affects morphological and physiological differentiation in *Streptomyces*. *Molecular Microbiology* **32**: 505-517.

69 **Debatisse, M. and F. Tolédo** (1995). Amplification génique, plasticité des génomes et oncogénèse. *Médecine/Sciences* **11**: 1099-1109.

70 **deHaseth, P. L. and J. D. Helmann** (1995). Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol Microbiol* **16**(5): 817-24.

71 **Dempsey, J. A. F., A. B. Wallace and J. G. Cannon** (1995). The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosomes of the two mapped strains of the closely related species *N.gonorrhoeae*. *Journal of Bacteriology* **177**: 6390-6400.

72 **Demuyter, P., D. Schneider, P. Leblond, J.-M. Simonet and B. Decaris** (1991). A chromosomal region as a hotspot for multiple rearrangements associated with genetic instability in *Streptomyces ambofaciens* DSM40697. *Journal of General Microbiology* **137**: 491-499.

- 73 **Deretic, V. and M. Konyecsni** (1990). A procaryotic regulatory factor with a histone H1-like carboxy-terminal domain: clonal variation of repeats within *algP*, a gene involved in regulation of mucoidy in *Pseudomonas aeruginosa*. *Journal of Bacteriology* **172**: 5544-5554.
- 74 **Deretic, V., M. J. Schurr, J. C. Boucher and D. W. Martin** (1994). Conversion of *Pseudomonas aeruginosa* to mucoidy in cystic fibrosis: environmental stress and regulation of bacterial virulence by alternative sigma factors. *J Bacteriol* **176**(10): 2773-80.
- 75 **Dombroski, A. J., W. A. Walter, M. T. Record, Jr., D. A. Siegele and C. A. Gross** (1992). Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. *Cell* **70**(3): 501-12.
- 76 **Duchene, A. M., C. J. Thompson and P. Mazodier** (1994). Transcriptional analysis of groEL genes in *Streptomyces coelicolor* A3(2). *Mol Gen Genet* **245**(1): 61-8.
- 77 **Dybvig, K.** (1993). DNA rearrangements and phenotypic switching in prokaryotes. *Molecular Microbiology* **10**(3): 465-471.
- 78 **Engelmann, S., C. Lindner and M. Hecker** (1995). Cloning, nucleotide sequence, and regulation of katE encoding a sigma B-dependent catalase in *Bacillus subtilis*. *J Bacteriol* **177**(19): 5598-605.
- 79 **Enz, S., V. Braun and J. H. Crosa** (1995). Transcription of the region encoding the ferric dicitrate-transport system in *Escherichia coli*: similarity between promoters for fecA and for extracytoplasmic function sigma factors. *Gene* **163**(1): 13-8.
- 80 **Erickson, J. W. and C. A. Gross** (1989). Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev* **3**(9): 1462-71.
- 81 **Evans, G. A., K. Lewis and B. E. Rothenberg** (1989). High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* **79**: 9-20.
- 82 **Farewell, A., K. Kvint and T. Nystrom** (1998). Negative regulation by RpoS: a case of sigma factor competition. *Mol Microbiol* **29**(4): 1039-51.
- 83 **Ferdows, M. S. and A. G. Barbour** (1989). Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi* the Lyme disease agent. *Proceedings of the National Academy of Sciences USA* **86**: 5969-5973.
- 84 **Fischer, G., B. Decaris and P. Leblond** (1997). Occurrence of deletions associated with genetic instability in *Streptomyces ambofaciens*, is independent of the linearity of the chromosomal DNA. *Journal of Bacteriology* **179**: 4553-4558.
- 85 **Fischer, G., A. C. Holl, J. N. Volff, D. Vandewiele, B. Decaris and P. Leblond** (1998a). Replication of the linear chromosomal DNA from the centrally located *oriC* of *Streptomyces ambofaciens* revealed by PFGE gene dosage analysis. *Research in Microbiology* **149**: 203-210.
- 86 **Fischer, G., T. Wenner, B. Decaris and P. Leblond** (1998b). Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proceedings of the National Academy of Sciences USA*: 14296-14301.
- 87 **Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan and J. Postlethwait** (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**(4): 1531-45.

- 88 **Fox, T. D. and J. Pero** (1974). New phage-SPO1-induced polypeptides associated with *Bacillus subtilis* RNA polymerase. *Proc Natl Acad Sci U S A* **71**(7): 2761-5.
- 89 **Fraser, C. M., S. Casjens, W. M. Huang, G. Sutton, R. Clayton, R. Lathigra, O. White, K. A. Ketchum, R. Dodson, E. Hickey, M. Gwinn, B. Dougherty, J. F. Tomb, R. Fleischmann, D. Richardson, J. Peterson, A. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. Van Vugt, N. Palmer, M. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Wadthay, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fujii, M. Cotton, K. Horst, K. Roberts, B. Hatch, O. Smith and J. C. Venter** (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**(6660): 580-6.
- 90 **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. L. Fuhrmann, D. T. Nguyen, T. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith and J. C. Venter** (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**(5235): 397-403.
- 91 **Freitas-Junior, L. H., E. Bottius, L. A. Pirrit, K. W. Deitsch, C. Scheidig, F. Guinet, U. Nehrbass, T. E. Wellems and A. Scherf** (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**(6807): 1018-22.
- 92 **French, S.** (1992). Consequences of replication fork movement through transcription units in vivo. *Science* **258**: 1362-1364.
- 93 **Fu, L. M. and C. S. Fu-Liu** (2002). Genome comparison of Mycobacterium tuberculosis and other bacteria. *OMICS* **6**: 199-206.
- 94 **Fujii, T., H. C. Gramajo, E. Takano and M. J. Bibb** (1996). redD and actII-ORF4, pathway-specific regulatory genes for antibiotic production in *Streptomyces coelicolor* A3(2), are transcribed in vitro by an RNA polymerase holoenzyme containing sigma hrdD. *J Bacteriol* **178**(11): 3402-5.
- 95 **Furano, A. V.** (1978). Direct demonstration of duplicate tuf genes in enteric bacteria. *Proc Natl Acad Sci U S A* **75**(7): 3104-8.
- 96 **Gallois, A.** (2003). "Polymorphisme intraspécifique des régions subtélomériques chez *Streptomyces ambifaciens*". DEA Microbiologie-Enzymologie-Nutrition
- 97 **Gehring, A. M., N. J. Yoo and R. Losick** (2001). RNA polymerase sigma factor that blocks morphological differentiation by *Streptomyces coelicolor*. *J Bacteriol* **183**(20): 5991-6.
- 98 **Gerth, U., E. Kruger, I. Derre, T. Msadek and M. Hecker** (1998). Stress induction of the *Bacillus subtilis* clpP gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Mol Microbiol* **28**(4): 787-802.
- 99 **Goffeau, A. e. a.** (1997). The yeast genome directory. *Nature* **387** (suppl.): 5.
- 100 **Goodman, M.** (1981a). Decoding the pattern of protein evolution. *Prog Biophys Mol Biol* **38**(2): 105-64.
- 101 **Goodman, M.** (1981b). Globin evolution was apparently very rapid in early vertebrates: a reasonable case against the rate-constancy hypothesis. *J Mol Evol* **17**(2): 114-20.
- 102 **Goodner, B., G. Hinkle, S. Gattung, N. Miller, M. Blanchard, B. Quorollo, B. S. Goldman, Y. Cao, M. Askenazi, C. Halling, L. Mullin, K. Houmiel, J. Gordon, M. Vaudin, O. Iartchouk, A. Epp, F. Liu, C. Wollam, M. Allinger, D. Doughty, C. Scott, C.**



- Lappas, B. Markelz, C. Flanagan, C. Crowell, J. Gurson, C. Lomo, C. Sear, G. Strub, C. Cielo and S. Slater (2001). Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**(5550): 2323-8.
- 103 Gorham, H. C., S. J. McGowan, P. R. Robson and D. A. Hodgson (1996). Light-induced carotenogenesis in *Myxococcus xanthus*: light-dependent membrane sequestration of ECF sigma factor CarQ by anti-sigma factor CarR. *Mol Microbiol* **19**(1): 171-86.
- 104 Gottesman, S. (1996). Proteases and their targets in *Escherichia coli*. *Annu Rev Genet* **30**: 465-506.
- 105 Gottesman, S. (1999). Regulation by proteolysis : developmental switches. *Curr. Opin. Microbiol.* **2**: 142-147.
- 106 Gottesman, S., E. Roche, Y. Zhou and R. T. Sauer (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes and Development* **12**: 1338-1347.
- 107 Gralla, J. D. and J. Collado-Vides (1996). Organization and function of transcription regulatory elements. *Escherichia coli and Salmonella cellular and molecular biology*. R. I. Curtiss III, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Schaechter, M., and Umberger, H.E. Washington, DC, American society for microbiology: 1232-1245.
- 108 Grandvalet, C., P. Servant and P. Mazodier (1997). Disruption of hspR, the repressor gene of the dnaK operon in *Streptomyces albus* G. *Mol Microbiol* **23**(1): 77-84.
- 109 Graves, M. C. and J. C. Rabinowitz (1986). In vivo and in vitro transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *J Biol Chem* **261**(24): 11409-15.
- 110 Gribskov, M. and R. R. Burgess (1986). Sigma factors from *E. coli*, *B. subtilis*, phage SP01, and phage T4 are homologous proteins. *Nucleic Acids Res* **14**(16): 6745-63.
- 111 Grossman, A. D. (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu Rev Genet* **29**: 477-508.
- 112 Gunes, G., B. Smith and P. Dyson (1999). Genetic instability associated with insertion of IS6100 into one end of the *Streptomyces lividans* chromosome. *Microbiology* **145 ( Pt 9)**: 2203-8.
- 113 Gust, B., G. L. Challis, K. Fowler, T. Kieser and K. F. Chater (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**(4): 1541-6.
- 114 Gust, B., T. Kieser and K. F. Chater (2002). Redirect technology: PCR-targeting system in *Streptomyces coelicolor*. Manual produced at the John Innes Centre.
- 115 Hackett, J. A., D. M. Feldser and C. W. Greider (2001). Telomere dysfunction increases mutation rate and genomic instability. *Cell* **106**(3): 275-86.
- 116 Haldenwang, W. G. (1995). The sigma factors of *Bacillus subtilis*. *Microbiol Rev* **59**(1): 1-30.
- 117 Hallenbeck, P. L., R. Lerchen, P. Hessler and S. Kaplan (1990). Roles of CfxA, CfxB, and external electron acceptors in regulation of ribulose 1,5-bisphosphate carboxylase/oxygenase expression in *Rhodobacter sphaeroides*. *J Bacteriol* **172**(4): 1736-48.

- 118 **Hanafusa, T. and H. Kinashi** (1992). The structure of an integrated copy of the giant linear plasmid SCP1 in the chromosome of *Streptomyces coelicolor* 2612. *Molecular and General Genetics* **231**: 363-368.
- 119 **Hanahan, D.** (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**(4): 557-80.
- 120 **Hanks, M., W. Wurst, L. Anson-Cartwright, A. B. Auerbach and A. L. Joyner** (1995). Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2. *Science* **269**(5224): 679-82.
- 121 **Harle, C., I. Kim, A. Angerer and V. Braun** (1995). Signal transfer through three compartments: transcription initiation of the *Escherichia coli* ferric citrate transport system from the cell surface. *Embo J* **14**(7): 1430-8.
- 122 **Harley, C. B. and R. P. Reynolds** (1987). Analysis of *E. coli* promoter sequences. *Nucleic Acids Res* **15**(5): 2343-61.
- 123 **Haselkorn, R.** (1992). Developmentally regulated gene rearrangements in prokaryotes. *Annual Review of Genetics* **26**: 113-130.
- 124 **Hass, R., S. Veit and T. F. Meyer** (1992). Silent pilin genes of *Neisseria gonorrhoeae* MS11 and the occurrence of related hypervariant sequences among other gonococcal isolates. *Molecular Microbiology* **6**: 197-208.
- 125 **Hecker, M., W. Schumann and U. Volker** (1996). Heat-shock and general stress response in *Bacillus subtilis*. *Mol Microbiol* **19**(3): 417-28.
- 126 **Hecker, M. and U. Volker** (1998). Non-specific, general and multiple stress resistance of growth-restricted *Bacillus subtilis* cells by the expression of the sigmaB regulon. *Mol Microbiol* **29**(5): 1129-36.
- 127 **Heichman, K. A. and R. C. Johnson** (1990). The hin invertasome: protein-mediated joining of distant recombination sites at the enhancer. *Science*(249): 511-517.
- 128 **Helmann, J. D.** (1995). Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res* **23**(13): 2351-60.
- 129 **Helmann, J. D.** (1999). Anti-sigma factors. *Curr Opin Microbiol* **2**(2): 135-41.
- 130 **Helmann, J. D. and M. J. Chamberlin** (1987). DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative sigma factor. *Proc Natl Acad Sci U S A* **84**(18): 6422-4.
- 131 **Helmann, J. D. and M. J. Chamberlin** (1988). Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839-72.
- 132 **Hengge-Aronis, R.** (1993). Survival of hunger and stress: the role of rpoS in early stationary phase gene regulation in *E. coli*. *Cell* **72**(2): 165-8.
- 133 **Hengge-Aronis, R.** (1996). Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Mol Microbiol* **21**(5): 887-93.
- 134 **Herman, C., D. Thevenet, P. Bouloc, W. G.C. and R. D'ari** (1998). Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Development* **12**: 1348-1355.

- 135 **Herman, C., D. Thevenet, R. D'Ari and P. Bouloc** (1995). Degradation of sigma 32, the heat shock regulator in *Escherichia coli*, is governed by HflB. *Proceedings of the National Academy of Sciences USA* **11**: 3516-3520.
- 136 **Higgins, D. G. and P. M. Sharp** (1988). CLUSTAL : a package for performing multiple sequence alignment on microcomputer. *Gene* **15**: 237-244.
- 137 **High, N. J., M. E. Deadman and E. R. Moxon** (1993). The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope  $\alpha$ Gal(1-4) $\beta$ Gal. *Molecular Microbiology* **9**: 1275-1282.
- 138 **Hill, T. M.** (1992). Arrest of bacterial DNA replication. *Annual Review of Microbiology* **46**: 603-633.
- 139 **Hirshfield, I. N., P. L. Bloch, R. A. Van Bogelen and F. C. Neidhardt** (1981). Multiple forms of lysyl-transfer ribonucleic acid synthetase in *Escherichia coli*. *J Bacteriol* **146**(1): 345-51.
- 140 **Hommes, N. G., L. A. Sayavedra-Soto and D. J. Arp** (1998). Mutagenesis and expression of amo, which codes for ammonia monooxygenase in *Nitrosomonas europaea*. *J Bacteriol* **180**(13): 3353-9.
- 141 **Hopwood, D. A.** (1967). Genetic analysis and genome structure in *Streptomyces coelicolor*. *Bacteriol Rev* **31**(4): 373-403.
- 142 **Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward and H. Schrempf** (1985). *Genetic Manipulation of Streptomyces-A Laboratory Manual*. Norwich, The John Innes Foundation.
- 143 **Hozumi, N. and S. Tonegawa** (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences USA* **73**: 3628-3632.
- 144 **Huang, C. H., Y. S. Lin, Y. A. Yang, S. W. Huang and C. W. Chen** (1998). The telomeres of *Streptomyces* chromosome contain conserved palindromic sequences with potential to form complex secondary structures. *Molecular Microbiology* **28**: 905-916.
- 145 **Hughes, M. K. and A. L. Hughes** (1993). Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol Biol Evol* **10**(6): 1360-9.
- 146 **Huisman, G. W. and R. Kolter** (1994). Sensing starvation: a homoserine lactone-dependent signaling pathway in *Escherichia coli*. *Science* **265**(5171): 537-9.
- 147 **Hunt, T. P. and B. Magasanik** (1985). Transcription of glnA by purified *Escherichia coli* components: core RNA polymerase and the products of glnF, glnG, and glnL. *Proc Natl Acad Sci U S A* **82**(24): 8453-7.
- 148 **Igarashi, K., N. Fujita and A. Ishihama** (1991). Identification of a subunit assembly domain in the alpha subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* **218**(1): 1-6.
- 149 **Ikeda, H., J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori and S. Omura** (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**(5): 526-31.
- 150 **Inouye, M. and S. Inouye** (1991). Retroelements in bacteria. *Trends Biochem Sci* **16**: 18-21.

- 151 **Inouye, S., M. Y. Hsu, S. Eagle and M. Inouye** (1989). Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell* **56**(4): 709-17.
- 152 **Ish-Horowicz, D. and J. F. Burke** (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Research* **9**: 2989-2998.
- 153 **Ishihama, A.** (1981). Subunit of assembly of Escherichia coli RNA polymerase. *Adv Biophys* **14**: 1-35.
- 154 **Jaskunas, S. R., L. Lindahl and M. Nomura** (1975). Identification of two copies of the gene for the elongation factor EF-Tu in E. coli. *Nature* **257**(5526): 458-62.
- 155 **JHong, H. J., M. S. Paget and M. J. Buttner** (2002). A signal transduction system in *Streptomyces coelicolor* that activates the expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. *Mol Microbiol* **44**(5): 1199-1211.
- 156 **Jishage, M. and A. Ishihama** (1998). A stationary phase protein in Escherichia coli with binding activity to the major sigma subunit of RNA polymerase. *Proc Natl Acad Sci U S A* **95**(9): 4953-8.
- 157 **Jones, G. H., M. S. Paget, L. Chamberlin and M. J. Buttner** (1997). Sigma-E is required for the production of the antibiotic actinomycin in *Streptomyces antibioticus*. *Molecular Microbiology* **23**: 169-178.
- 158 **Jumas-Bilak, E., S. Michaux-Charanchon, G. Bourg, D. O'Callaghan and M. Ramuz** (1998). Differences in chromosome number and genome rearrangements in the genus *Brucella*. *Molecular Microbiology* **27**: 99-107.
- 159 **Kalkus, J., C. Dörrrie, D. Fischer, M. Reh and H. G. Schlegel** (1993). The giant linear plasmid pHG207 from *Rhodococcus* sp. encoding hydrogen autotrophy: characterization of the plasmid and its termini. *Journal of General Microbiology* **139**: 2055-2065.
- 160 **Kalkus, J., M. Reh and H. G. Schlegel** (1990). Hydrogen autotrophy of *Nocardia opaca* strains is encoded by linear megaplasmids. *J Gen Microbiol* **136** ( Pt 6): 1145-51.
- 161 **Kang, J. G., M. Y. Hahn, A. Ishihama and J. H. Roe** (1997). Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from *Streptomyces coelicolor* A3(2). *Nucleic Acids Res* **25**(13): 2566-73.
- 162 **Kang, J. G., M. S. Paget, Y. J. Seok, M. Y. Hahn, J. B. Bae, J. S. Hahn, C. Kleanthous, M. J. Buttner and J. H. Roe** (1999). RsrA, an anti-sigma factor regulated by redox change. *Embo J* **18**(15): 4292-8.
- 163 **Kassavetis, G. A. and E. P. Geiduschek** (1984). Defining a bacteriophage T4 late promoter: bacteriophage T4 gene 55 protein suffices for directing late promoter recognition. *Proc Natl Acad Sci U S A* **81**(16): 5101-5.
- 164 **Kataoka, M., T. Seki and T. Yoshida** (1991). Five genes involved in self-transmission of pSN22, a *Streptomyces* plasmid. *J Bacteriol* **173**(13): 4220-8.
- 165 **Kato, F., T. Hino, A. Nakaji, M. Tanaka and Y. Koyama** (1995). Carotenoid synthesis in *Streptomyces setonii* ISP5395 is induced by the gene crtS, whose product is similar to a sigma factor. *Mol Gen Genet* **247**(3): 387-90.
- 166 **Kawakami, K., K. Ito and Y. Nakamura** (1992). Differential regulation of two genes encoding lysyl-tRNA synthetases in Escherichia coli: lysU-constitutive mutations compensate for a lysS null mutation. *Mol Microbiol* **6**(13): 1739-45.

- 167 **Keener, J. and S. Kustu** (1988). Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *Proc Natl Acad Sci U S A* **85**(14): 4976-80.
- 168 **Keener, J. and M. Nomura** (1993). Dominant lethal phenotype of a mutation in the -35 recognition region of Escherichia coli sigma 70. *Proc Natl Acad Sci U S A* **90**(5): 1751-5.
- 169 **Kelemen, G. H., G. L. Brown, J. Kormanec, L. Potuckova, K. F. Chater and M. J. Buttner** (1996). The positions of the sigma-factor genes, whiG and sigF, in the hierarchy controlling the development of spore chains in the aerial hyphae of Streptomyces coelicolor A3(2). *Mol Microbiol* **21**(3): 593-603.
- 170 **Kelemen, G. H., P. H. Viollier, J. Tenor, L. Marri, M. J. Buttner and C. J. Thompson** (2001). A connection between stress and development in the multicellular prokaryote Streptomyces coelicolor A3(2). *Mol Microbiol* **40**(4): 804-14.
- 171 **Kieser, H. M., T. Kieser and D. A. Hopwood** (1992). A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. *Journal of Bacteriology* **174**: 5496-5507.
- 172 **Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood** (2000). *Practical Streptomyces genetics*. Norwich, The John Innes Foundation.
- 173 **Kim, S. and T. Komano** (1997). The plasmid R64 thin pilus identified as a type IV pilus. *Journal of Bacteriology* **179**: 3594-3603.
- 174 **Kitten, T. and A. G. Barbour** (1990). Juxtaposition of expressed variable antigen genes with a conserved telomere in the bacterium *Borrelia hermsii*. *Proceedings of the National Academy of Sciences USA* **87**: 6077-6081.
- 175 **Kitten, T., A. V. Barrera and A. G. Barbour** (1993). Intragenic recombination and a chimeric outer membrane protein in the relapsing fever agent *Borrelia hermsii*. *Journal of Bacteriology* **175**: 2516-2522.
- 176 **Klar, A. J. S.** (1987). Determination of the yeast cell lineage. *Cell* **49**: 433-435.
- 177 **Klintworth, R., M. Husemann, J. Salnikow and B. Bowien** (1985). Chromosomal and plasmid locations for phosphoribulokinase genes in Alcaligenes eutrophus. *J Bacteriol* **164**(2): 954-6.
- 178 **Klose, K. E. and J. J. Mekalanos** (1998). Distinct roles of an alternative sigma factor during both free-swimming and colonizing phases of the Vibrio cholerae pathogenic cycle. *Mol Microbiol* **28**(3): 501-20.
- 179 **Kolb, A., D. Kotlarz, S. Kusano and A. Ishihama** (1995). Selectivity of the Escherichia coli RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res* **23**(5): 819-26.
- 180 **Kolsto, A. B.** (1997). Dynamic bacterial genome organization. *Molecular Microbiology* **21**(2): 241-248.
- 181 **Koonin, E. V., A. R. Mushegian, M. Y. Galperin and D. R. Walker** (1997). Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol Microbiol* **25**(4): 619-37.
- 182 **Kormanec, J., D. Homerova, I. Barak and B. Sevcikova** (1999). A new gene, sigG, encoding a putative alternative sigma factor of *Streptomyces coelicolor* A3(2). *FEMS Microbiology Letters* **172**: 153-158.

- 183 **Kormanec, J. and B. Sevcikova** (2002a). The stress-response sigma factor sigma(H) controls the expression of ssgB, a homologue of the sporulation-specific cell division gene ssgA, in *Streptomyces coelicolor* A3(2). *Mol Genet Genomics* **267**(4): 536-43.
- 184 **Kormanec, J. and B. Sevcikova** (2002b). Stress-response sigma factor sigma(H) directs expression of the gltB gene encoding glutamate synthase in *Streptomyces coelicolor* A3(2). *Biochim Biophys Acta* **1577**(1): 149-54.
- 185 **Kormanec, J., B. Sevcikova, N. Halgasova, R. Knirschova and B. Rezuchova** (2000). Identification and transcriptional characterization of the gene encoding the stress-response sigma factor sigma(H) in *Streptomyces coelicolor* A3(2). *FEMS Microbiol Lett* **189**(1): 31-8.
- 186 **Krawiec, S. and M. Riley** (1990). Organization of the bacterial chromosome. *Microbiological Reviews* **54**: 502-539.
- 187 **Kroos, L. and Y. T. Yu** (2000). Regulation of sigma factor activity during *Bacillus subtilis* development. *Curr Opin Microbiol* **3**(6): 553-60.
- 188 **Kross, L., B. Zhang, H. Ichikawa and Y. N. Yu** (1999). Control of  $\sigma$  factor activity during *Bacillus subtilis* sporulation. *Molecular Microbiology* **31**: 1285-1294.
- 189 **Krum, J. G. and S. A. Ensign** (2001). Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2. *J Bacteriol* **183**(7): 2172-7.
- 190 **Kuhstoss, S. and R. N. Rao** (1991). Analysis of the integration function of the streptomycete bacteriophage phi C31. *J Mol Biol* **222**(4): 897-908.
- 191 **Kulasekara, H. D. and I. C. Blomfield** (1999). The molecular basis for the specificity of *fimE* in the phase variation of type 1 fimbriae of *Escherichia coli* K-12. *Molecular Microbiology* **31**: 1171-1181.
- 192 **Kumar, A., R. A. Malloch, N. Fujita, D. A. Smillie, A. Ishihama and R. S. Hayward** (1993). The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J Mol Biol* **232**(2): 406-18.
- 193 **Kunkel, B., R. Losick and P. Stragier** (1990). The *Bacillus subtilis* gene for the development transcription factor sigma K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. *Genes and Development* **4**: 525-535.
- 194 **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Düsterhöft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, A. Fabret, E. Ferrari, B. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, J. Golightly, E. Grandi, F. Guiseppi, J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, B. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R. P. Mellado, M. Mizuno, A. Moestl, S. Nakai, M. Noback, B. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, M. Prescott, E. Presecan, P. Pujic, A. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A.**

Sekowska, S. J. Seror, P. Serror, B. S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra and A. Tognoni (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**(6657): 249-56.

195 Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi and K. Hiramatsu (2001). Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* **357**(9264): 1225-40.

196 Kutsukake, K. and T. Iino (1994). Role of the FliA-FlgM regulatory system on the transcriptional control of the flagellar regulon and flagellar formation in *Salmonella typhimurium*. *J Bacteriol* **176**(12): 3598-605.

197 Kutsukake, K., S. Iyoda, K. Ohnishi and T. Iino (1994). Genetic and molecular analyses of the interaction between the flagellum-specific sigma and anti-sigma factors in *Salmonella typhimurium*. *Embo J* **13**(19): 4568-76.

198 Lammers, P. J., J. W. Golden and R. Haselkorn (1986). Identification and sequence of a gene required for a developmentally regulated DNA excision in *Anabaena*. *Cell* **44**: 905-911.

199 Lampson, B. C., J. Sun, M. Y. Hsu, J. Vallejo-Ramirez, S. Inouye and M. Inouye (1989). Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* **243**(4894 Pt 1): 1033-8.

200 Lange, R. and R. Hengge-Aronis (1991). Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol Microbiol* **5**(1): 49-59.

201 Lange, R. and R. Hengge-Aronis (1994). The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev* **8**(13): 1600-12.

202 Leblond, P. and B. Decaris (1994). New insights into the genetic instability of *Streptomyces*. *FEMS Microbiology Letters* **123**: 225-232.

203 Leblond, P. and B. Decaris (1999). 'Unstable' linear chromosomes : the case of *Streptomyces*. *Organization of the Prokaryotic Genome*. R. L. Charlebois. Washington, D.C., American Society for Microbiology: 235-261.

204 Leblond, P., P. Demuyter, J. M. Simonet and B. Decaris (1991). Genetic instability and associated genome plasticity in *Streptomyces ambofaciens* : pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. *Journal of Bacteriology* **173**: 4229-4233.

205 Leblond, P., G. Fischer, F. X. Francou, F. Berger, M. Guerineau and B. Decaris (1996). The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol Microbiol* **19**(2): 261-71.

206 Leblond, P., M. Redenbach and J. Cullum (1993). Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). *Journal of Bacteriology* **175**: 3422-3429.

207 Lee, H. S., Y. Ohnishi and S. Horinouchi (2001). A sigmaB-like factor responsible for carotenoid biosynthesis in *Streptomyces griseus*. *J Mol Microbiol Biotechnol* **3**(1): 95-101.



- 208 **Lesley, S. A. and R. R. Burgess** (1989). Characterization of the *Escherichia coli* transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. *Biochemistry* **28**(19): 7728-34.
- 209 **Leveque, F., M. Gazeau, M. Fromant, S. Blanquet and P. Plateau** (1991). Control of *Escherichia coli* lysyl-tRNA synthetase expression by anaerobiosis. *J Bacteriol* **173**(24): 7903-10.
- 210 **Lezhava, A., T. Mizukami, T. Kajitani, D. Kameoka, M. Redenbach, H. Shinkawa, O. Nimi and H. Kinashi** (1995a). Physical map of the linear chromosome of *Streptomyces griseus*. *J Bacteriol* **177**(22): 6492-8.
- 211 **Lezhava, A., T. Mizukami, T. Kajitani, D. Kameoka, M. Redenbach, H. Shinkawa, O. Nimi and H. Kinashi** (1995b). Physical map of the linear chromosome of *Streptomyces griseus*. *Journal of Bacteriology* **177**: 6492-6498.
- 212 **Li, L. and C. K. Lim** (2000). A novel large plasmid carrying multiple beta-lactam resistance genes isolated from a *Klebsiella pneumoniae* strain. *J Appl Microbiol* **88**(6): 1038-48.
- 213 **Li, W. H.** (1985). Accelerated evolution following gene duplication and its implication for the neutralist-selectionist controversy. *Population genetics and molecular evolution*. Berlin, Springer-Verlag: 333-352.
- 214 **Li, X. and M. Noll** (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions. *Nature* **367**(6458): 83-7.
- 215 **Lin, Y. S., H. M. Kieser, D. A. Hopwood and C. W. Chen** (1993). The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Molecular Microbiology* **10**: 923-933.
- 216 **Linardopoulou, E., H. C. Mefford, O. Nguyen, C. Friedman, G. van den Engh, D. G. Farwell, M. Coltrera and B. J. Trask** (2001). Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum Mol Genet* **10**(21): 2373-83.
- 217 **Liu, J., W. M. Cosby and P. Zuber** (1999). Role of lon and ClpX in the post-translational regulation of a sigma subunit of RNA polymerase required for cellular differentiation in *Bacillus subtilis*. *Mol Microbiol* **33**(2): 415-28.
- 218 **Liu, S.-L. and K. E. Sanderson** (1996). Highly plastic chromosomal organization in *Salmonella typhi*. *Proceedings of the National Academy of Sciences USA* **93**: 10303-10308.
- 219 **Llorente, B., P. Durrens, A. Malpertuy, M. Aigle, F. Artiguenave, G. Blandin, M. Bolotin-Fukuhara, E. Bon, P. Brottier, S. Casaregola, B. Dujon, J. de Montigny, A. Lepingle, C. Neueglise, O. Ozier-Kalogeropoulos, S. Potier, W. Saurin, F. Tekaia, C. Toffano-Nioche, M. Wesolowski-Louvel, P. Wincker, J. Weissenbach, J. Souciet and C. Gaillardin** (2000a). Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett* **487**(1): 122-33.
- 220 **Llorente, B., A. Malpertuy, C. Neueglise, J. de Montigny, M. Aigle, F. Artiguenave, G. Blandin, M. Bolotin-Fukuhara, E. Bon, P. Brottier, S. Casaregola, P. Durrens, C. Gaillardin, A. Lepingle, O. Ozier-Kalogeropoulos, S. Potier, W. Saurin, F. Tekaia, C. Toffano-Nioche, M. Wesolowski-Louvel, P. Wincker, J. Weissenbach, J. Souciet and B. Dujon** (2000b). Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Lett* **487**(1): 101-12.
- 221 **Lonetto, M., M. Gribskov and C. A. Gross** (1992). The  $\sigma$ 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology* **174**: 3843-3849.



- 222 **Lonetto, M. A., K. L. Brown, K. E. Rudd and M. J. Buttner** (1994). Analysis of the *Streptomyces coelicolor* sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91**(16): 7573-7.
- 223 **Lorenz, M. G. and W. Wackernagel** (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* **58**(3): 563-602.
- 224 **MacNeil, D. J.** (1988). Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. *J Bacteriol* **170**(12): 5607-12.
- 225 **Mager, W. H. and A. J. De Kruijff** (1995). Stress-induced transcriptional activation. *Microbiol Rev* **59**(3): 506-31.
- 226 **Malhotra, A., E. Severinova and S. A. Darst** (1996). Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell* **87**(1): 127-36.
- 227 **Marck, C.** (1988). 'DNA Strider' : a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Research* **11**: 1829-1836.
- 228 **Martinez, A. and R. Kolter** (1997). Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J Bacteriol* **179**(16): 5188-94.
- 229 **Masterson, J.** (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421-424.
- 230 **McClain, M. S., I. C. Blomfield and B. I. Eisenstein** (1991). Roles of *fimB* and *fimE* in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. *Journal of Bacteriology* **173**: 5308-5314.
- 231 **Mendes, M. V., J. F. Aparicio and J. F. Martin** (2000). Complete nucleotide sequence and characterization of pSNA1 from pimarinin-producing *Streptomyces natalensis* that replicates by a rolling circle mechanism. *Plasmid* **43**(2): 159-65.
- 232 **Merrick, M. J.** (1993). In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol* **10**(5): 903-9.
- 233 **Mira, A., H. Ochman and N. A. Moran** (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**(10): 589-96.
- 234 **Missiakas, D., M. P. Mayer, M. Lemaire, C. Georgopoulos and S. Raina** (1997). Modulation of the *Escherichia coli* sigmaE (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol Microbiol* **24**(2): 355-71.
- 235 **Missiakas, D. and S. Raina** (1998). The extracytoplasmic function sigma factors : role and regulation. *Molecular Microbiology* **28**: 1059-1066.
- 236 **Muffler, A., M. Barth, C. Marschall and R. Hengge-Aronis** (1997). Heat shock regulation of sigmaS turnover: a role for DnaK and relationship between stress responses mediated by sigmaS and sigma32 in *Escherichia coli*. *J Bacteriol* **179**(2): 445-52.
- 237 **Muller, H. J.** (1930). Types of visible variations induced by X-rays in *Drosophila*. *Journal of Genetics* **22**: 299-334.
- 238 **Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich** (1986). Specific enzymatic amplification of DNA in vitro : the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51**: 263-273.

- 239 **Murphy, K. C.** (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**(8): 2063-71.
- 240 **Mushegian, A. R. and E. V. Koonin** (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* **93**(19): 10268-73.
- 241 **Musialowski, M. S., F. Flett, G. B. Scott, G. Hobbs, C. Smith and S. G. Oliver** (1994). Functional evidence that the principal DNA replication origin of the *Streptomyces coelicolor* chromosome is close to the dnaA-gyrB region. *Journal of Bacteriology* **176**: 5123-5125.
- 242 **Ng, I., S. L. Liu and K. E. Sanderson** (1999). Role of genomic rearrangements in producing new ribotypes of *Salmonella typhi*. *Journal of Bacteriology* **181**: 3536-3541.
- 243 **NG, I., S.-L. Liu and K. E. Sanderson** (1999). Role of genomic rearrangements in producing new ribotypes of *Salmonella typhi*. *Journal of Bacteriology* **181**: 3536-3541.
- 244 **Nowak, M. A., M. C. Boerlijst, J. Cooke and J. M. Smith** (1997). Evolution of genetic redundancy. *Nature* **388**(6638): 167-71.
- 245 **Ochman, H., J. G. Lawrence and E. A. Groisman** (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**(6784): 299-304.
- 246 **Ochs, M., S. Veitinger, I. Kim, D. Welz, A. Angerer and V. Braun** (1995). Regulation of citrate-dependent iron transport of *Escherichia coli*: fecR is required for transcription activation by FecI. *Mol Microbiol* **15**(1): 119-32.
- 247 **O'Connor, T. J., P. Kanellis and J. R. Nodwell** (2002). The ramC gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Mol Microbiol* **45**(1): 45-57.
- 248 **Ohno, S.** (1970). *Evolution by gene duplication*. London.
- 249 **Ohta, T.** (1988). Time for acquiring a new gene by duplication. *Proc Natl Acad Sci U S A* **85**(10): 3509-12.
- 250 **Okanishi, M., K. Suzuki and H. Umezawa** (1974). Formation and reversion of Streptomyces protoplasts: cultural condition and morphological study. *J Gen Microbiol* **80**(2): 389-400.
- 251 **Omura, S., H. Ikeda, J. Ishikawa, A. Hanamoto, C. Takahashi, M. Shinose, Y. Takahashi, H. Horikawa, H. Nakazawa, T. Osonoe, H. Kikuchi, T. Shiba, Y. Sakaki and M. Hattori** (2001). Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A* **98**(21): 12215-20.
- 252 **Ovchinnikov, Y. A., V. M. Lipkin, N. N. Modyanov, O. Y. Chertov and Y. V. Smirnov** (1977). Primary structure of alpha-subunit of DNA-dependent RNA polymerase from *Escherichia coli*. *FEBS Lett* **76**(1): 108-11.
- 253 **Ovchinnikov, Y. A., G. S. Monastyrskaya, V. V. Gubanov, S. O. Guryev, O. Chertov, N. N. Modyanov, V. A. Grinkevich, I. A. Makarova, T. V. Marchenko, I. N. Polovnikova, V. M. Lipkin and E. D. Sverdlov** (1981). The primary structure of *Escherichia coli* RNA polymerase. Nucleotide sequence of the rpoB gene and amino-acid sequence of the beta-subunit. *Eur J Biochem* **116**(3): 621-9.
- 254 **Ovchinnikov, Y. A., G. S. Monastyrskaya, V. V. Gubanov, S. O. Guryev, I. S. Salomatina, T. M. Shuvaeva, V. M. Lipkin and E. D. Sverdlov** (1982). The primary structure

of E. coli RNA polymerase, Nucleotide sequence of the rpoC gene and amino acid sequence of the beta'-subunit. *Nucleic Acids Res* **10**(13): 4035-44.

255 **Paget, M. S., J. B. Bae, M. Y. Hahn, W. Li, C. Kleanthous, J. H. Roe and M. J. Buttner** (2001a). Mutational analysis of RstA, a zinc-binding anti-sigma factor with a thiol-disulphide redox switch. *Mol Microbiol* **39**(4): 1036-47.

256 **Paget, M. S., L. Chamberlin, A. Atrih, S. J. Foster and M. J. Buttner** (1999). Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**(1): 204-11.

257 **Paget, M. S., J. G. Kang, J. H. Roe and M. J. Buttner** (1998). sigmaR, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *Embo J* **17**(19): 5776-82.

258 **Paget, M. S., V. Molle, G. Cohen, Y. Aharonowitz and M. J. Buttner** (2001b). Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol Microbiol* **42**(4): 1007-20.

259 **Pandza, K., G. Pfalzer, J. Cullum and D. Hranueli** (1997). Physical mapping shows that the unstable oxytetracycline gene cluster of *Streptomyces rimosus* lies close the ends of the linear chromosome. *Microbiology* **143**: 1493-1501.

260 **Pandza, S., G. Biukovic, A. Paravic, A. Dadbin, J. Cullum and D. Hranueli** (1998). Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. *Molecular Microbiology* **28**: 1165-1176.

261 **Pang, X., X. Zhou, Y. Sun and Z. Deng** (2002). Physical map of the linear chromosome of *Streptomyces hygroscopicus* 10-22 deduced by analysis of overlapping large chromosomal deletions. *J Bacteriol* **184**(7): 1958-65.

262 **Pansegrau, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich and C. M. Thomas** (1994). Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol* **239**(5): 623-63.

263 **Papadakis, M. N. and G. P. Patrinos** (1999). Contribution of gene conversion in the evolution of the human beta-like globin gene family. *Hum Genet* **104**(2): 117-25.

264 **Petersohn, A., M. Brigulla, S. Haas, J. D. Hoheisel, U. Volker and M. Hecker** (2001). Global analysis of the general stress response of *Bacillus subtilis*. *J Bacteriol* **183**(19): 5617-31.

265 **Picardeau, M. and V. Vincet** (1997). Characterization of large linear plasmids in Mycobacteria. *Journal of Bacteriology* **179**: 2753-2756.

266 **Polo, S., O. Guerini, M. Sosio and G. Deho** (1998). Identification of two linear plasmids in the actinomycete *Planobispora rosea*. *Microbiology* **144** ( Pt 10): 2819-25.

267 **Polyakov, A., E. Severinova and S. A. Darst** (1995). Three-dimensional structure of E. coli core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell* **83**(3): 365-73.

268 **Popham, D. L. and P. Stragier** (1992). Binding of the *Bacillus subtilis* spoIVCA product to the recombination sites of the element interrupting the  $\sigma^k$ -encoding gene. *Proceedings of the National Academy of Sciences USA* **89**: 5991-5995.

- 269 Potuckova, L., G. H. Kelemen, K. C. Findlay, M. A. Lonetto, M. J. Buttner and J. Kormanec (1995). A new RNA polymerase sigma factor, sigma F, is required for the late stages of morphological differentiation in *Streptomyces* spp. *Mol Microbiol* **17**(1): 37-48.
- 270 Powers, P. A. and O. Smithies (1986). Short gene conversions in the human fetal globin gene region: a by-product of chromosome pairing during meiosis? *Genetics* **112**(2): 343-58.
- 271 Pridham, T. G., P. Anderson, C. Foley, L. A. Lidenfelser and R. G. Benedict (1957a). A selection of media for maintenance and taxonomic study of *Streptomyces*. *Antibiot. Ann.* **1956-1957**: 947-953.
- 272 Pridham, T. G., P. Anderson, C. Foley, L. A. Lidenfelser and R. G. Benedict (1957b). A selection of medium for maintenance and taxonomic study of *Streptomyces*. *Antibiot Ann* **1956/57**: 947-953.
- 273 Pryde, F. E., H. C. Gorham and E. J. Louis (1997). Chromosome ends: all the same under their caps. *Current Opinion in Genetics and Development* **7**: 822-828.
- 274 Qin, Z. and S. N. Cohen (2002). Survival mechanisms for *Streptomyces* linear replicons after telomere damage. *Mol Microbiol* **45**(3): 785-94.
- 275 Ramos-Onsins, S. and M. Aguade (1998). Molecular evolution of the Cecropin multigene family in *Drosophila*. Functional genes vs. pseudogenes. *Genetics* **150**(1): 157-71.
- 276 Redenbach, M., F. Flett, W. Piendl, I. Glocker, U. Rauland, O. Wafzig, R. Kliem, P. Leblond and J. Cullum (1993). The *Streptomyces lividans* 66 chromosome contains a 1 MB deletogenic region flanked by two amplifiable regions. *Molecular and General Genetics* **241**: 255-262.
- 277 Redenbach, M., H. M. Kieser, D. Denapaite, A. Eichner, J. Cullum and D. A. Hopwood (1996). A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Molecular Microbiology* **21**: 77-96.
- 278 Redenbach, M., J. Scheel and U. Schmidt (2000). Chromosome topology and genome size of selected actinomycetes species. *Antonie Van Leeuwenboek* **78**(3-4): 227-35.
- 279 Reeves, A. R., D. A. Post and T. J. Vanden Boom (1998). Physical-genetic map of the erythromycin-producing organism *Saccharopolyspora erythraea*. *Microbiology* **144** ( Pt 8): 2151-9.
- 280 Restrepo, B. I., T. Kitten, C. J. Carter, D. Infante and A. G. Barbour (1992). Subtelomeric expression regions of *Borrelia hermsii* linear plasmids are highly polymorphic. *Molecular Microbiology* **6**: 3299-3311.
- 281 Rezuchova, B., I. Barak and J. Kormanec (1997). Disruption of a sigma factor gene, *sigF*, affects an intermediate stage of spore pigment production in *Streptomyces aureofaciens*. *FEMS Microbiology Letters* **153**: 371-377.
- 282 Rhodes, L. D., A. M. Coady and M. S. Strom (2002). Expression of duplicate *msa* genes in the salmonid pathogen *Renibacterium salmoninarum*. *Appl Environ Microbiol* **68**(11): 5480-7.
- 283 Riley, M. and A. Anilionis (1978). Evolution of the bacterial genome. *Annu Rev Microbiol* **32**: 519-60.
- 284 Rosa, P. A., T. Schwan and D. Hogan (1992). Recombination between genes encoding major outer surface proteins A and B of *Borrelia burgdorferi*. *Molecular Microbiology* **20**: 3031-3040.

- 285 **Rubio, J. P., J. K. Thompson and A. F. Cowman** (1996). The *var* genes of *plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *EMBO Journal* **15**: 4069-4077.
- 286 **Rudner, R., A. Murray and N. Huda** (1999). Is there a link between mutation rates and the stringent response in *Bacillus subtilis*? *Ann N Y Acad Sci* **870**: 418-22.
- 287 **Ryding, N. J., G. H. Kelemen, C. A. Whatling, K. Flardh, M. J. Buttner and K. F. Chater** (1998). A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **29**(1): 343-57.
- 288 **S., G.** (1999). Regulation by proteolysis : developmental switches. *Current Opinion in Microbiology* **2**: 142-147.
- 289 **Sabelnikov, A. G., B. Greenberg and S. A. Lacks** (1995). An extended -10 promoter alone directs transcription of the DpnII operon of *Streptococcus pneumoniae*. *J Mol Biol* **250**(2): 144-55.
- 290 **Sambrook, J., E. F. Fritsch and T. Maniatis** (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New-York, Cold Spring Harbor Laboratory Press.
- 291 **Sanger, F., S. Nicklen and A. R. Coulson** (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**: 5463-5467.
- 292 **SantaLucia, J.** (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences USA* **95**: 1460-1465.
- 293 **Sasse-Dwight, S. and J. D. Gralla** (1988). Probing the *Escherichia coli* *glnALG* upstream activation mechanism in vivo. *Proc Natl Acad Sci U S A* **85**(23): 8934-8.
- 294 **Schauner, C., A. Dary, A. Lebrhi, P. Leblond, B. Decaris and P. Germain** (1999). Modulation of lipid metabolism and spiramycin biosynthesis in *Streptomyces ambofaciens* unstable mutants. *Applied in Environmental Microbiology* **65**: 2730-2737.
- 295 **Schneider, D., e. B. Aigl, P. Leblond, J. M. Simonet and B. Decaris** (1993). Analysis of genome instability in *Streptomyces ambofaciens*. *Journal of General Microbiology* **139**: 2559-2567.
- 296 **Scott, T. N. and M. I. Simon** (1982). Genetic analysis of the mechanism of the *Salmonella* phase variation site specific recombination system. *Molecular and General Genetics* **188**: 313-321.
- 297 **Sela, S., D. Yogeve, S. Razin and H. Bercovier** (1989). Duplication of the *tuf* gene: a new insight into the phylogeny of eubacteria. *J Bacteriol* **171**(1): 581-4.
- 298 **Sen, R. and D. Dasgupta** (1996). Conformational changes of *E. coli* RNA polymerase during transcription initiation. *Biophys Chem* **57**(2-3): 269-78.
- 299 **Seoighe, C. and K. H. Wolfe** (1999). Updated map of duplicated regions in the yeast genome. *Gene* **238**(1): 253-61.
- 300 **Servin-Gonzalez, L., A. I. Sampieri, J. Cabello, L. Galvan, V. Juarez and C. Castro** (1995). Sequence and functional analysis of the *Streptomyces phaeochromogenes* plasmid pJV1 reveals a modular organization of *Streptomyces* plasmids that replicate by rolling circle. *Microbiology* **141** ( Pt 10): 2499-510.

- 301 **Sevcikova, B., O. Benada, O. Kofronova and J. Kormanec** (2001). Stress-response sigma factor sigma(H) is essential for morphological differentiation of *Streptomyces coelicolor* A3(2). *Arch Microbiol* **177**(1): 98-106.
- 302 **Sevcikova, B. and J. Kormanec** (2002). Activity of the *Streptomyces coelicolor* stress-response sigma factor sigmaH is regulated by an anti-sigma factor. *FEMS Microbiol Lett* **209**(2): 229-35.
- 303 **Severinov, K., D. Fenyo, E. Severinova, A. Mustaev, B. T. Chait, A. Goldfarb and S. A. Darst** (1994). The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *J Biol Chem* **269**(33): 20826-8.
- 304 **Shapiro, J. A.** (1969). Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *Journal of Molecular Biology* **28**: 93-105.
- 305 **Sharp, P. M.** (1991). Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J Mol Evol* **33**(1): 23-33.
- 306 **Shen, S. H., J. L. Slightom and O. Smithies** (1981). A history of the human fetal globin gene duplication. *Cell* **26**(2 Pt 2): 191-203.
- 307 **Shimeld, S. M.** (1999). Gene function, gene networks and the fate of duplicated genes. *Semin Cell Dev Biol* **10**(5): 549-53.
- 308 **Shorenstein, R. G. and R. Losick** (1973). Purification and properties of the sigma subunit of ribonucleic acid polymerase from vegetative *Bacillus subtilis*. *J Biol Chem* **248**(17): 6163-9.
- 309 **Sidow, A.** (1996). Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev* **6**(6): 715-22.
- 310 **Skeidsvoll, J. and P. M. Ueland** (1995). Analysis of double-stranded DNA by capillary electrophoresis with laser-induced fluorescence detection using the monomeric dye SYBR green I. *Anal Biochem* **231**(2): 359-65.
- 311 **Slightom, J. L., A. E. Blechl and O. Smithies** (1980). Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**(3): 627-38.
- 312 **Souciet, J., M. Aigle, F. Artiguenave, G. Blandin, M. Bolotin-Fukuhara, E. Bon, P. Brottier, S. Casaregola, J. de Montigny, B. Dujon, P. Durrens, C. Gaillardin, A. Lepingle, B. Llorente, A. Malpertuy, C. Neugeglise, O. Ozier-Kalogeropoulos, S. Potier, W. Saurin, F. Tekaia, C. Toffano-Nioche, M. Wesolowski-Louvel, P. Wincker and J. Weissenbach** (2000). Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* **487**(1): 3-12.
- 313 **Spradling, A. C. and A. P. Mahowald** (1980). Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA* **77**: 1096-1100.
- 314 **Stolyar, S., A. M. Costello, T. L. Peeples and M. E. Lidstrom** (1999). Role of multiple gene copies in particulate methane monooxygenase activity in the methane-oxidizing bacterium *Methylococcus capsulatus* Bath. *Microbiology* **145** ( Pt 5): 1235-44.
- 315 **Stragier, P., C. Bonamy and C. Karmazyn-Campelli** (1988). Processing of a sporulation sigma factor in *Bacillus subtilis*: how morphological structure could control gene expression. *Cell* **52**(5): 697-704.

- 316 **Stragier, P. and R. Losick** (1996). Molecular genetics of sporulation in *Bacillus subtilis*. *Annu Rev Genet* **30**: 297-41.
- 317 **Strohl, W. R.** (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Research* **20**(5): 961-974.
- 318 **Sun, J., G. H. Kelemen, J. M. Fernandez-Abalos and M. J. Bibb** (1999). Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). *Microbiology* **145** ( Pt 9): 2221-7.
- 319 **Szabo, M., D. Maskell, P. Butler, J. Love and R. Moxon** (1992). Use of the chromosomal gene fusions to investigate the role of repetitive DNA in regulation of genes involved in lipopolysaccharide biosynthesis in *Haemophilus influenzae*. *Journal of Bacteriology* **174**: 7245-7252.
- 320 **Takeuchi, T., H. Sawada, F. Tanaka and I. Matsuda** (1996). Phylogenetic analysis of *Streptomyces* spp causing potato scab based on 16S rRNA sequences. *International Journal of Systematic Bacteriology* **46**: 476-479.
- 321 **Tan, H. and K. F. Chater** (1993). Two developmentally controlled promoters of *Streptomyces coelicolor* A(3)2 that resemble the major class of motility-related promoters in other bacteria. *Journal of Bacteriology* **175**: 933-940.
- 322 **Tanaka, K., S. Kusano, N. Fujita, A. Ishihama and H. Takahashi** (1995). Promoter determinants for *Escherichia coli* RNA polymerase holoenzyme containing sigma 38 (the rpoS gene product). *Nucleic Acids Res* **23**(5): 827-34.
- 323 **Tanaka, K., T. Shiina and H. Takahashi** (1988). Multiple principal sigma factor homologs in eubacteria : identification of the 'rpoD box'. *Science* **18**: 1040-1042.
- 324 **Tauch, A., I. Homann, S. Mormann, S. Ruberg, A. Billault, B. Bathe, S. Brand, O. Brockmann-Gretza, C. Ruckert, N. Schischka, C. Wrenger, J. Hoheisel, B. Mockel, K. Huthmacher, W. Pfefferle, A. Puhler and J. Kalinowski** (2002). Strategy to sequence the genome of *Corynebacterium glutamicum* ATCC 13032: use of a cosmid and a bacterial artificial chromosome library. *J Biotechnol* **95**(1): 25-38.
- 325 **Thompson, C. J., J. M. Ward and D. A. Hopwood** (1980). DNA cloning in *Streptomyces*: resistance genes from antibiotic-producing species. *Nature* **286**(5772): 525-7.
- 326 **Tonegawa, S.** (1976). Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. *Proceedings of the National Academy of Sciences USA* **73**: 203-207.
- 327 **Uyeno, T. and G. R. Smith** (1972). Tetraploid origin of the karyotype of *catostomid* fishes. *Science* **175**(22): 644-6.
- 328 **Van Deutekom, J. C., E. Bakker, R. J. Lemmers, M. J. van der Wielen, E. Bik, M. H. Hofker, G. W. Padberg and R. R. Frants** (1996). Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetics counselling and etiology of FSHD1. *Human Molecular Genetics* **5**: 1997-2003.
- 329 **Viala, J., G. Rapoport and P. Mazodier** (2000). The clpP multigenic family in *Streptomyces lividans*: conditional expression of the clpP3 clpP4 operon is controlled by PopR, a novel transcriptional activator. *Mol Microbiol* **38**(3): 602-12.

- 330 **Vijgenboom, E., L. P. Woudt, P. W. Heinstra, K. Rietveld, J. van Haarlem, G. P. van Wezel, S. Shochat and L. Bosch** (1994). Three tuf-like genes in the kirromycin producer *Streptomyces ramocissimus*. *Microbiology* **140** ( Pt 4): 983-98.
- 331 **Viollier, P. H., G. H. Kelemen, G. E. Dale, K. T. Nguyen, M. J. Buttner and C. J. Thompson** (2003a). Specialized osmotic stress response systems involve multiple SigB-like sigma factors in *Streptomyces coelicolor*. *Mol Microbiol* **47**(3): 699-714.
- 332 **Viollier, P. H., A. Weihofen, M. Folcher and C. J. Thompson** (2003b). Post-transcriptional regulation of the *Streptomyces coelicolor* stress responsive sigma factor, SigH, involves translational control, proteolytic processing, and an anti-sigma factor homolog. *J Mol Biol* **325**(4): 637-49.
- 333 **Volff, J. N. and J. Altenbuchner** (2000). A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. *FEMS Microbiol Lett* **186**(2): 143-50.
- 334 **Volff, J. N., D. Vandewiele, J. M. Simonet and B. Decaris** (1993). Ultraviolet light, mitomycin C and nitrous acid induce genetic instability in *Streptomyces ambofaciens* ATCC23877. *Mutation Research* **287**: 141-156.
- 335 **Volker, U., S. Engelmann, B. Maul, S. Riethdorf, A. Volker, R. Schmid, H. Mach and M. Hecker** (1994). Analysis of the induction of general stress proteins of *Bacillus subtilis*. *Microbiology* **140** ( Pt 4): 741-52.
- 336 **Wada, M. and Y. Nakamura** (1996). Unique telomeric expression site of major-surface-glycoprotein genes of *Pneumocystis carinii*. *DNA Res* **3**(2): 55-64.
- 337 **Wagner, A.** (2001). Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends Genet* **17**(5): 237-9.
- 338 **Walsh, J. B.** (1995). How often do duplicated genes evolve new functions? *Genetics* **139**(1): 421-8.
- 339 **Ward, J. M., G. R. Janssen, T. Kieser, M. J. Bibb and M. J. Buttner** (1986). Construction and characterisation of a series of multi-copy promoter-probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase gene from Tn5 as indicator. *Molecular and General Genetics* **203**: 468-478.
- 340 **Wenner, T.** (1998). Translocation d'extrémité chromosomique chez *Streptomyces ambofaciens* : recombinaison homologue entre copies d'un gène putatif codant un facteur  $\sigma$ . France, Université Henri Poincaré, Nancy 1.
- 341 **Wenner, T., V. Roth, B. Decaris and P. Leblond** (2002). Intragenomic and intraspecific polymorphism of the 16S-23S rDNA internally transcribed sequences of *Streptomyces ambofaciens*. *Microbiology* **148**(Pt 3): 633-42.
- 342 **Wilkie, A. O., D. R. Higgs, K. A. Rack, V. J. Buckle, N. K. Spurr, N. Fischel-Ghodsian, I. Ceccherini, W. R. Brown and P. C. Harris** (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64**(3): 595-606.
- 343 **Willems, H., C. Jager and G. Baljer** (1998). Physical and genetic map of the obligate intracellular bacterium *Coxiella burnetii*. *J Bacteriol* **180**(15): 3816-22.
- 344 **Willems, R., A. Raul, H. G. J. van der Heide, A. R. ter Avest and F. R. Mooi** (1990). Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *The EMBO Journal* **9**: 2803-2809.



- 345 Willey, J., J. Schwedock and R. Losick (1993). Multiple extracellular signals govern the production of a morphogenetic protein involved in aerial mycelium formation by *Streptomyces coelicolor*. *Genes Dev* 7(5): 895-903.
- 346 Wilson, C. and A. J. Dombroski (1997). Region 1 of sigma70 is required for efficient isomerization and initiation of transcription by Escherichia coli RNA polymerase. *J Mol Biol* 267(1): 60-74.
- 347 Wise, A. A. and C. W. Price (1995). Four additional genes in the sigB operon of Bacillus subtilis that control activity of the general stress factor sigma B in response to environmental signals. *J Bacteriol* 177(1): 123-33.
- 348 Wolf, S. G., D. Frenkiel, T. Arad, S. E. Finkel, R. Kolter and A. Minsky (1999). DNA protection by stress-induced biocrystallization. *Nature* 400(6739): 83-5.
- 349 Wolfe, K. H. and D. C. Shields (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634): 708-13.
- 350 Wood, D. W., J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, V. K. Okura, Y. Zhou, L. Chen, G. E. Wood, N. F. Almeida, Jr., L. Woo, Y. Chen, I. T. Paulsen, J. A. Eisen, P. D. Karp, D. Bovee, Sr., P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutuyavin, R. Levy, M. J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z. Y. Zhao, M. Dolan, F. Chumley, S. V. Tingey, J. F. Tomb, M. P. Gordon, M. V. Olson and E. W. Nester (2001). The genome of the natural genetic engineer Agrobacterium tumefaciens C58. *Science* 294(5550): 2317-23.
- 351 Wood, V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, B. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. Mclean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, B. Odell, K. Oliver, S. O'neil, D. Pearson, M. A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, D. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, C. Fritz, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. Del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerrutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, C. Ussery, B. G. Barrell and P. Nurse (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415(6874): 871-80.
- 352 Wosten, M. M. (1998). Eubacterial sigma-factors. *FEMS Microbiol Rev* 22(3): 127-50.
- 353 Wren, B. W. (1991). A family of clostridial and Streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. *Molecular Microbiology* 5: 797-803.
- 354 Yamanaka, K., L. Fang and M. Inouye (1998). The CspA family in Escherichia coli: multiple gene duplication for stress adaptation. *Mol Microbiol* 27(2): 247-55.

- 355 Yang, C. C., C. H. Huang, C. Y. Li, Y. G. Tsay, S. C. Lee and C. W. Chen (2002). The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. *Mol Microbiol* **43**(2): 297-305.
- 356 Yin, Y., M. A. Tainsky, F. Z. Bischoff, L. C. Strong and G. M. Wahl (1992). Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell* **18**: 937-948.
- 357 Yura, T., H. Nagai and H. Mori (1993). Regulation of the heat-shock response in Bacteria. *Annual Review of Genetics* **47**: 321-350.
- 358 Zenitani, H., K. Goshi, K. Higashiyama, T. Uchida, K. Hiratsu, D. Kameoka, A. Lezhava, H. Shinkawa and H. Kinashi (1999). Dynamic structural changes of the linear chromosome of *Streptomyces griseus*. 11th International Symposium on the Biology of the Actinomycetes, Crete.
- 359 Zhang, B., P. Struffi and L. Kroos (1999). sigmaK can negatively regulate sigE expression by two different mechanisms during sporulation of *Bacillus subtilis*. *J Bacteriol* **181**(13): 4081-8.
- 360 Zhou, X., Z. Deng, J. L. Firmin, D. A. Hopwood and T. Kieser (1988). Site-specific degradation of *Streptomyces lividans* DNA during electrophoresis in buffers contaminated with ferrous iron. *Nucleic Acids Research* **25**: 4341-4352.

## ❧ PUBLICATIONS ❧

**PUBLICATION N°1**

**Intragenomic and intraspecific polymorphism of the  
16S-23S rDNA internally transcribed sequences of  
*Streptomyces ambofaciens***

Thomas Wenner, **Virginie Roth**, Bernard Decaris et Pierre Leblond

*Microbiology, Vol.148, pp 633-642*

## Intragenomic and intraspecific polymorphism of the 16S–23S rDNA internally transcribed sequences of *Streptomyces ambofaciens*

Thomas Wenner, Virginie Roth, Bernard Decaris and Pierre Leblond

Author for correspondence: Pierre Leblond. Tel.: +33 03 83 91 22 04. Fax: +33 03 83 91 25 00.  
e-mail: leblond@nancy.inra.fr

Laboratoire de Génétique et Microbiologie UMR INRA 1128-IFR110, Faculté des Sciences, Université Henri Poincaré – Nancy 1, BP 239, Boulevard des Aiguillettes, F-54506 Vandœuvre-lès-Nancy, France

The nucleotide composition of the internally transcribed sequences (ITSs) of the six rDNA operons of two strains of *Streptomyces ambofaciens* were determined. Four variable and five conserved nucleotide blocks were distinguished. Five different modular organizations were revealed for each strain and no homologous loci showed the same succession of blocks. This suggests that recombination frequently occurs between the rDNA loci, leading to the exchange of nucleotide blocks. The modular structure was also observed within the ITSs of *Streptomyces coelicolor* M145, which is closely related to *Streptomyces ambofaciens*, and *Streptomyces griseus* 2247, showing the same number of constant blocks but with fewer variable regions. This confirms that a high degree of ITS variability is a common characteristic among *Streptomyces* spp. The functional significance of the combinations of variable and constant nucleotide blocks of the ITS was examined by *in silico* prediction of secondary structures from nucleotide sequences. The secondary structures were shown to be analogous whatever the combination of variable/constant blocks at the intragenomic, intraspecific and interspecific levels.

Keywords: rDNA evolution, gene conversion, recombination, intergenic spacer, ITS

### INTRODUCTION

*Streptomyces* are high-GC, Gram-positive soil bacteria, belonging to the order *Actinomycetales* within the class *Actinobacteria*. They exhibit a complex life cycle, undergoing differentiation leading to the production of numerous economically important secondary metabolites, including antibiotics. They possess linear chromosomal DNA of about 8 Mb, with a high GC content (70–74 mol %). The extremities of their DNA are subject to large rearrangements, encompassing about 25% of the total genome size (for a review see Leblond & Decaris, 1999). Six copies of rDNA are present in all of the *Streptomyces* spp. that have been studied so far [including *Streptomyces ambofaciens* (Berger *et al.*, 1996; Cole & Saint-Girons, 1999)], with the exception of *Streptomyces venezuelae* for which seven operons

were characterized (La Farina *et al.*, 1996). These *rrn* operons share the typical bacterial organization, 16S–23S–5S (Cole & Saint-Girons, 1999), with no tRNA genes found to be present in the 16S–23S spacers in any of the *Actinomycetales* that have been studied so far.

While 16S rRNA sequence analysis is a powerful tool for inferring inter- or intrageneric relationships, due to the strong conservation of its nucleotide sequence across species and genera, the 16S–23S internally transcribed sequence (ITS) region, which shows a faster rate of evolution, provides information concerning intraspecific relationships. Thus, the use of ITSs in phylogenetic studies is limited to detecting recently diverged species and to the typing of bacteria (for a review see Gürtler & Stanisich, 1996).

Different levels of variability within ITSs have been described at the intraspecific level. In most cases, ITS variability results from nucleotide substitutions or from the presence or absence of blocks of nucleotides, as has been observed in *Escherichia coli* (Anton *et al.*, 1998, 1999), *Salmonella enterica* (Luz *et al.*, 1998) and lactobacilli (Nour, 1998). In some cases this variation has been used to infer phylogenetic relationships, such

**Abbreviations:** ATCC, American Type Culture Collection; DSM, Deutsche Sammlung von Mikroorganismen und Zellkulturen; dsPS, double-stranded processing site; ITS, internally transcribed sequence; c, constant region; v, variable region; vc, variable/constant region.

The GenBank accession numbers for the sequences reported in this paper can be found in Table 1.

as in the genus *Bifidobacterium* where strains of the same species share greater than 93 % nucleotide identity (Leblond-Bourget *et al.*, 1996). These phylogenetic relationships were consistent with those determined by 16S rRNA sequence analysis.

The finding that every bacterial strain has a unique ITS allows for the rapid identification and typing of bacteria. The homogeneity of ITSs at the intragenomic level has been empirically revealed by the ability to directly determine the nucleotide sequence of PCR products derived from the genomic DNA, as shown for enterococci, streptococci, lactobacilli (Naimi *et al.*, 1997; Nour *et al.*, 1995) and some members of the *Actinobacteria* [e.g. *Bifidobacterium* spp. (Leblond-Bourget *et al.*, 1996) and *Tropheryma whippelii* (Maiwald *et al.*, 2000)]. Sequencing of whole sets of *rrn* loci has also demonstrated this intragenomic homogeneity. Data for these sets are available for organisms whose genomes have been completely sequenced, or where *rrn* loci have been specifically cloned and analysed (Gürtler, 1999, and references therein). For example, in *Brucella melitensis* the three rRNA operons are identical over the entire length of the *rrn* loci (Bricker, 2000). When present, intragenomic variability consists of a few substitutions over the entire ITS length or, more frequently, of the presence or absence of tRNA genes (Christensen *et al.*, 2000; Gürtler *et al.*, 1999; Itehan *et al.*, 2000; Naimi *et al.*, 1997). Apart from tRNA insertions, ITS homogeneity was thought to be a universal feature and it was proposed that frequent gene conversion events between *rrn* loci might be responsible for the strong homogeneity.

Several cases have been reported where a high level of intragenomic ITS variability has been observed. In *Vibrio cholerae* (Lan & Reeves, 1998), *Vibrio mimicus* (Chun *et al.*, 1999), *Staphylococcus aureus* (Gürtler, 1999) and *Haemophilus parainfluenzae* (Privitera *et al.*, 1998), the ITS consists of a mosaic of variable regions with intervening constant nucleotide blocks. The maintenance of such heterogeneity can result from the same mechanism responsible for homogeneity in other species, i.e. gene conversion can result in an exchange of blocks of nucleotides to generate mosaic structures (for a review see Gürtler, 1999). In *Streptomyces*, typing of *Streptomyces albidoflavus* strains based on the number, size and sequence of cloned ITSs has led to the description of such a modular structure (Hain *et al.*, 1997).

The cloning and analysis of the ITSs of the complete set of *rrn* loci for two *Streptomyces ambofaciens* strains has provided an opportunity to study the structure and evolutionary mechanisms of these sequences. Comparison of these ITSs with those of other *Streptomyces* species (*Streptomyces coelicolor* M145 and *Streptomyces griseus* 2247) has allowed the determination of the different types of variability, i.e. variability in the number, size and sequence of the variable and constant blocks. Here we propose a mechanism which may produce this variability. The use of ITSs in phylogenetic analysis is also evaluated. We also show that the significance of the different nucleotide block arrange-

ments is related to the secondary structure of the transcribed spacer and a possible functional role in RNA maturation is proposed.

## METHODS

**Organisms, reagents and enzymes.** The bacterial strains used in this work were *Streptomyces ambofaciens* ATCC 23877 (Pinnert-Sindico *et al.*, 1955) and *Escherichia coli* SURE (Stratagene). Restriction enzymes and molecular biology reagents were purchased from New England Biolabs and Roche Diagnostics. Cosmids of *Streptomyces ambofaciens* DSM 40697 bearing rDNA loci were previously isolated by Berger *et al.* (1996).

**DNA extractions and gene library construction.** Total DNA for the preparation of the *Streptomyces ambofaciens* gene library was extracted as described in Kieser *et al.* (2000). Cosmid DNA was extracted from *Escherichia coli* by the alkaline lysis method (Sambrook *et al.*, 1989).

The *Streptomyces ambofaciens* ATCC 23877 gene library was constructed from partial *Bam*HI digests of total DNA in Supercos1 (ampicillin-, kanamycin-resistant; Stratagene) using the Gigapack III XL Packaging Extracts (Stratagene), according to the manufacturer's instructions.

**DNA labelling and hybridization.** The cosmid DNA and PCR products were labelled using a non-radioactive digoxigenin-labelling kit (DIG DNA Labelling and Detection Kit; Roche Diagnostics), as directed by the supplier. DNA fragments were transferred onto nylon membranes (Hybond-N; Amersham Pharmacia Biotech) and were cross-linked by exposure to UV light (Sambrook *et al.*, 1989). Colony transfer onto Nytran-N membranes (Schleicher & Schuell) was carried out according to Sambrook *et al.* (1989). Hybridizations with digoxigenin-labelled probes were carried out as directed by the supplier (Roche Diagnostics) at 68 °C. The signals were detected with light emission and acquisition of results was done with a Fluor-S Multimager (Bio-Rad).

**Amplification of DNA and purification of PCR products.** Total DNA from *Streptomyces ambofaciens* ATCC 23877 and cosmids containing rDNA loci from *Streptomyces ambofaciens* DSM 40697 were digested with *Bam*HI, purified by a phenol/chloroform extraction and then used as the template for PCR amplification of the ITS. The PCR primers (MWG Biotech) used were L (5'-GCTGGATCACCTCCTTTCT-3', 3' end of the 16S rDNA) and R (5'-CTGGTGCCAAGGCATCCA-3', 5' end of the 23S rDNA) (Berger *et al.*, 1996). PCR was performed as described by the suppliers (Appligene Oncor, Roche Diagnostics) in a Perkin Elmer thermal cycler. Approximately 10 ng DNA was subjected to a PCR consisting of an initial denaturation step at 95 °C for 10 min, followed by 30 reaction cycles consisting of denaturation at 95 °C for 1 min, annealing at 63 °C for 1 min and extension at 72 °C for 2 min, followed by a final extension step at 72 °C for 7 min. The PCR products were electrophoresed in a 2% agarose gel and further purified using the High Pure PCR product Purification Kit (Roche Diagnostics) for nucleotide sequencing.

**DNA sequencing and analysis.** Nucleotide sequences of the PCR products were determined using the ABI Prism Dye Terminator Cycle Sequencing Kit (Perkin Elmer), with the L and R primers, and an ABI 310 Prism automated DNA sequencer (Applied Biosystems). Nucleotide sequences were aligned using the CLUSTAL W program (Thompson *et al.*, 1994). Secondary structures were sought with the MFOLD program

**Table 1.** ITS GenBank accession numbers

The ITS data for *Streptomyces ambofaciens* DSM 40697 and ATCC 23877 were derived during this study, with the exception of ATCC 23877 rDNA-D (this study and Pernodet *et al.*, 1989). *Streptomyces coelicolor* M145 sequence data were retrieved from the *Streptomyces coelicolor* sequencing project (Sanger Centre, [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)). *Streptomyces griseus* data were from a direct submission to GenBank by H. Shinkawa and K. Inada.

Species/strain and ITS locus	Accession no.
<i>Streptomyces ambofaciens</i>	
DSM 40697	
rDNA-A	AY027675
rDNA-B	AY027676
rDNA-C	AY027677
rDNA-D	AY027678
rDNA-E	AY027679
rDNA-F	AY027680
<i>Streptomyces ambofaciens</i>	
ATCC 23877	
rDNA-A	AY027681
rDNA-B	AY027682
rDNA-C	AY027683
rDNA-D	M27245
	AY027684
rDNA-E	AY027685
rDNA-F	AY027686
<i>Streptomyces coelicolor</i>	
M145	
rDNA-A	AL356612
rDNA-B	AL109848
rDNA-C	AL353861
rDNA-D	AL353862
rDNA-E	AL031031
rDNA-F	AL079345
<i>Streptomyces griseus</i> 2247	
rDNA-A	AB030567
rDNA-B	AB030568
rDNA-C	AB030569
rDNA-D	AB030570
rDNA-D'	AB030571
rDNA-E	AB030572

version 3.1 (Mathews *et al.*, 1999; <http://bioinfo.math.rpi.edu/~mfold/rna/form1.cgi>). The accession numbers for the nucleotide sequences determined or used in this study are listed in Table 1.

## RESULTS

### Sequencing of the *Streptomyces ambofaciens* rDNA loci

To identify clones for the six rDNA loci of *Streptomyces ambofaciens* ATCC 23877, about 2000 recombinant clones from a cosmid library were screened using a

probe consisting of the 16S–23S ITS, obtained by the amplification of total DNA with the short universal oligonucleotides L and R (see Methods). Eighteen clones were selected. Restriction enzyme analysis and hybridization with the 16S–23S ITS DNA probe allowed six groups of overlapping recombinant cosmids to be distinguished, with a mean insert size of 40 kb. The rDNA loci were contained in *Bam*HI fragments of 9, 11 (two distinct loci), 13 (two distinct loci) and 16 kb. These sizes were fully consistent with those estimated from the hybridization of the same probe against a Southern blot of *Bam*HI-digested chromosomal DNA of *Streptomyces ambofaciens* ATCC 23877 (Pernodet *et al.*, 1989). A representative recombinant cosmid for each of the six cloned *rrn* loci was chosen for further experiments.

DNA from each of the six cosmids was used as a probe in Southern hybridization experiments with the *Bam*HI digests of cosmids, including the six *Streptomyces ambofaciens* DSM 40697 rDNA loci (data not shown). This hybridization allowed us to form pairs of homologous *rrn* loci for the two *Streptomyces ambofaciens* strains, in spite of the restriction polymorphism. The rDNA loci were then labelled A–F according to the nomenclature used for the genetic maps of *Streptomyces ambofaciens* DSM 40697 (Berger *et al.*, 1996; Leblond *et al.*, 1996) and *Streptomyces coelicolor* M145 (Redenbach *et al.*, 1996).

PCR amplifications, using the primers L and R, were performed on the recombinant cosmid DNA for each rDNA locus of both of the *Streptomyces ambofaciens* strains. The size of the ITS PCR products was estimated by electrophoresis to be about 300 bp for all 12 loci, consistent with the presence of a single PCR product derived from the total genomic DNA using the same pair of primers (data not shown).

### The 16S–23S spacer of *Streptomyces ambofaciens* is a mosaic of variable and constant nucleotide blocks

The nucleotide sequence of each PCR product revealed a polymorphism in the size of the ITSs. The ITSs varied from 299 bp for rDNA-C (ITS of *rrnC* locus) to 306 bp for rDNA-A of *Streptomyces ambofaciens* ATCC 23877, and from 300 bp for rDNA-B/-C/-E to 304 bp for rDNA-A of *Streptomyces ambofaciens* DSM 40697 (Fig. 1). This was in good agreement with the sizes estimated on agarose gels (data not shown).

The alignment of the *Streptomyces ambofaciens* ITSs is shown in Fig. 1. Five constant regions (c1–c5) and four variable regions (v1–v4) could be distinguished. Constant regions shared 100% nucleotide identity and were 13, 67, 21, 62 and 49 nt long for c1, c2, c3, c4 and c5, respectively. Analysis of the variable loci revealed three haplotypes each for v1–v3 (v1.1, v1.2 and v1.3, v2.1, v2.2 and v2.3, and v3.1, v3.2 and v3.3, respectively) and two haplotypes for v4 (v4.1 and v4.2) (Fig. 1). Polymorphisms within a variable region, consisting of one or two substitutions (double-underlined in Fig. 1), were







**Fig. 2.** Schematic comparison of the ITSs of *Streptomyces griseus* 2247, *Streptomyces ambofaciens* DSM 40697, *Streptomyces ambofaciens* ATCC 23877 and *Streptomyces coelicolor* M145. The ITS designations are as defined in Fig. 1. In *Streptomyces griseus*, D' was omitted from the alignment (see text). Regions specific to *Streptomyces griseus* are black and those specific to *Streptomyces coelicolor* are light grey. Regions variable between species, but which are constant within the same species, are designated by vc and represented by horizontal hatched lines in *Streptomyces griseus*, by diagonal hatched lines in *Streptomyces ambofaciens* and by hatched grey in *Streptomyces coelicolor*. The nucleotide numbering was derived from the alignment. Nucleotide differences in the constant region between *Streptomyces griseus* and the other strains are shown, whereas the positions that differ between variable regions are not.

natural transformation has not been observed in *Streptomyces* spp. (Rhodes, 1986).

#### Intraspecific comparison shows evidence for recombination between rDNA loci

The two *Streptomyces ambofaciens* strains examined here shared the same pool of variable regions, with no strain-specific nucleotide blocks. Furthermore, homologous loci exhibited different variable region combinations. In fact, ITS A from DSM 40697 and ATCC 23877 shared only v1.1, v1 and v4 for ITS B, v4 for ITS C, v1 and v4 for ITS D, v1 for ITS E and v4 for ITS F. In contrast, non-homologous rDNA loci could be identical (e.g. ITS B and F in DSM 40697 were identical to ITS F and E in ATCC 23877, respectively). In all, eight different combinations were observed.

Thus, a given ITS corresponds to a mosaic of variable blocks. This specific structure may result from recombination between rDNA loci. Such events are known to trigger large chromosomal rearrangements, as was observed in *Salmonella typhi* (Liu & Sanderson, 1996) and in *Brucella suis* (Jumas-Bilak *et al.*, 1998). The six

cosmids of *Streptomyces ambofaciens* ATCC 23877, bearing the rDNA loci and its flanking sequences, were hybridized onto the set of *Streptomyces ambofaciens* DSM 40697 rDNA-containing cosmids. The *Bam*HI fragments harbouring the *rrn* locus were revealed in all six patterns and with all six DNA probes. In addition, the rDNA flanking regions of the *Streptomyces ambofaciens* ATCC 23877 cosmid revealed sequences belonging to their *Streptomyces ambofaciens* DSM 40697 counterpart. These data support the conservation of the genetic map between the two *Streptomyces ambofaciens* strains and they rule out the formation of large DNA rearrangements involving rDNA loci since the divergence from their last common ancestor. Therefore, in *Streptomyces ambofaciens* rDNA recombination may result in shuffling of the variable regions between rDNA loci.

#### Interspecific ITS comparison

A comparison of the ITSs of *Streptomyces ambofaciens* with those of *Streptomyces coelicolor* M145, to which it is closely related, and with those of *Streptomyces griseus* 2247 was performed (Fig. 2). *Streptomyces coelicolor*

M145 possesses three different sequences, ITS A, ITS B/C/D and ITS E/F, with only four constant regions and three variable regions (v2–v4). The absence of v1 may result from a DNA rearrangement affecting c1, v1 and c2 and a further loss of variability due to recombination. This hypothesis is supported by the shorter size of c1 and c2 in *Streptomyces coelicolor*, as compared to both of the *Streptomyces ambofaciens* strains and to *Streptomyces griseus* 2247. The remaining part of c2, and the entire c3 and c5 regions of *Streptomyces coelicolor* are identical to the corresponding region in *Streptomyces ambofaciens*.

Identity between c4 sequences of *Streptomyces ambofaciens* and *Streptomyces coelicolor* is disrupted by a sequence of 16 bp and by a sequence of 18 bp, respectively (positions 205–221 and 175–193, respectively). This constitutes a constant region at the species level and was named the variable/constant region (vc) (Fig. 2). In addition, some of the variable sequences (v2.4/v3.4) are specific to *Streptomyces coelicolor*. The modular structure arrangement described for *Streptomyces ambofaciens* was also present in *Streptomyces coelicolor* M145 (e.g. ITS A exhibits v2 and v3 of ITS E/F and v4 of ITS B/C/D).

*Streptomyces griseus* possesses three different types of ITSs, namely ITS A/B/C, ITS D/E and ITS D'. The latter is highly divergent at the nucleotide level from the other five, preventing the detection of nucleotide blocks by sequence comparison. Thus, it was not considered in the following analyses and was not reported in Fig. 2. Two variable regions only, v2 and v3, could be distinguished. The absence of v1 and v4 could be explained by the loss of variability resulting from replacement of the whole set of rDNA loci by a single combination of nucleotide blocks specific of *Streptomyces griseus* (Fig. 2).

The sequences of the variable regions define new types of v3 within *Streptomyces griseus* (v3.5 and v3.6), shorter and with less than 59% sequence identity with the v3 sequences of *Streptomyces ambofaciens*. The sequences of v2 regions are similar to v2.1 and v2.2 of *Streptomyces ambofaciens*, with six differences for v2.1 and four for v2.2 (not shown on Fig. 2). Nine nucleotides differ between *Streptomyces griseus* and *Streptomyces ambofaciens* in the constant regions (reported in Figs 2 and 3). Furthermore, vc, defined by the comparison between *Streptomyces ambofaciens* and *Streptomyces coelicolor*, is also found in the ITSs of *Streptomyces griseus* (positions 195–212). However, Hain *et al.* (1997) studied 10 ITSs of two strains of the same species, *Streptomyces albidoflavus*. Comparison of these sequences revealed six constant and five variable regions. The v4 and v5 sequences described by Hain *et al.* (1997) correspond to the vc and to v4 of *Streptomyces ambofaciens*, respectively. Thus, whereas the vc is common for *Streptomyces ambofaciens* strains, it corresponds to a variable region in *Streptomyces albidoflavus*.

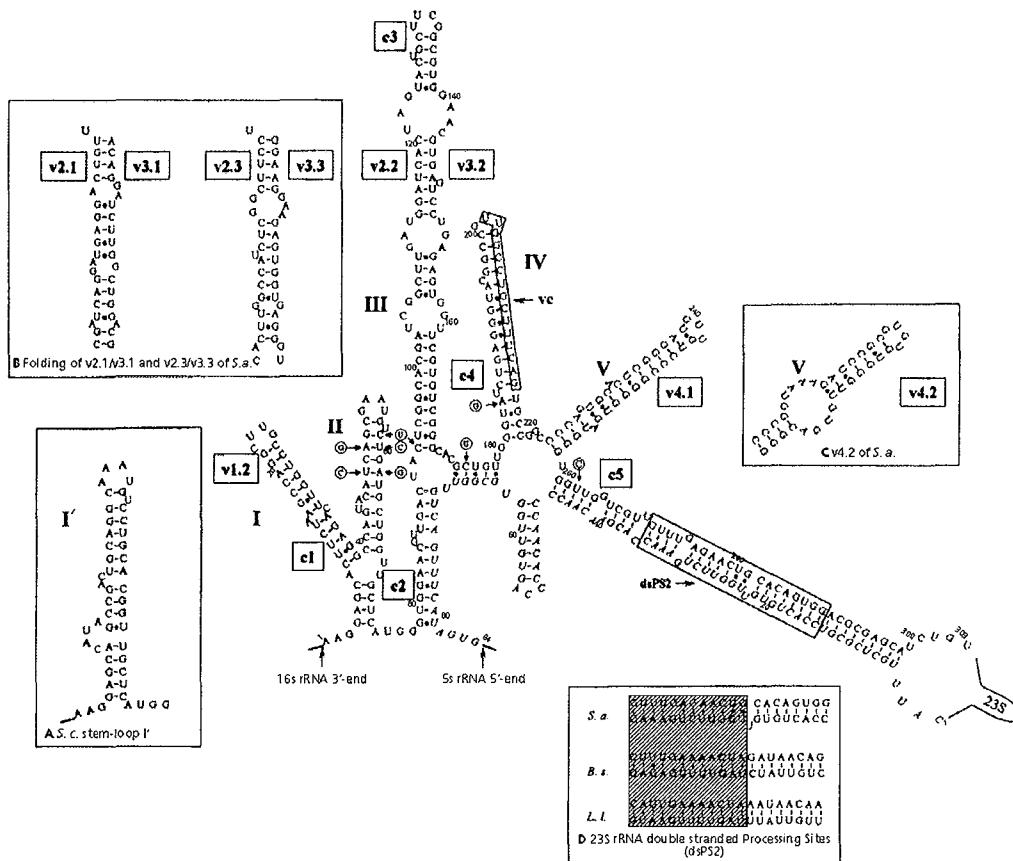
Recombination has been proposed to explain the rearrangement of rDNA operons, notably the rearrange-

ment of 20–50 nt sequence blocks within the ITS of *Haemophilus parainfluenzae* (Privitera *et al.*, 1998). Reciprocal recombination between rDNA loci could also be responsible for the formation of large chromosomal DNA rearrangements, as observed in *Salmonella typhi* (Liu & Sanderson, 1996). No large chromosomal rearrangement was observed between the genetic maps of *Streptomyces ambofaciens* DSM 40697 and ATCC 23877. Thus, recombination leads to the shuffling of the variable region between the rDNA loci. Moreover, comparison between *Streptomyces ambofaciens*, *Streptomyces coelicolor* and *Streptomyces griseus* suggests that recombination between rDNA loci can lead to the shuffling of nucleotide blocks and can either maintain variability or result in homogeneous ITSs.

Reciprocal and non-reciprocal recombination could be involved in this phenomenon. Large chromosomal rearrangements would result from the occurrence of a single crossing-over event between rDNA loci. To explain the shuffling of variable blocks, double crossing-over events, with at least one event (initiated or ending) within the ITSs, should be envisaged. In contrast, crossing-over events occurring on both sides of the ITS would lead to the total exchange of the ITS and would not generate its mosaic structure. The double crossing-over events that would involve the 16S and 23S gene sequences could explain the existence of identical ITSs at non-homologous loci (for example, rDNA-B of DSM 40697 and rDNA-F of ATCC 23877). A single recombination event, with no exchange of flanking sequences (non-reciprocal recombination or gene conversion), is perhaps more likely to lead to the shuffling of DNA blocks. This mechanism may be acting to make the variable blocks homogeneous, as observed in *Streptomyces coelicolor* and in *Streptomyces griseus*.

### Structural analysis

The functional significance of ITS variability was considered at the level of secondary structure interactions. Secondary structure models were constructed using the MPOLD program (version 3.1; Mathews *et al.*, 1999) to examine the interactions within the 16S–23S spacer for the eight *Streptomyces ambofaciens* combinations and between the 16S–23S and 23S–5S spacers of rRNA-D of *Streptomyces ambofaciens* ATCC 23877 and the six rRNAs of *Streptomyces coelicolor* M145 and *Streptomyces griseus* 2247. [Secondary structures, resulting from the folding of the 16S–23S spacer onto the 23S–5S spacer, are thought to participate in the maturation process of the pre-rRNA by RNase III in *Escherichia coli* (reviewed by Apirion & Miczak, 1993).] These theoretical models predicted the presence of long hairpins and a few small bubbles, with a free energy of about  $-180 \text{ kcal mol}^{-1}$  at 30 °C. Moreover, these structures were analogous whatever the sequence used. As shown in Fig. 3, constant as well as variable nucleotide blocks are involved in stem-loop secondary structures. Thus, v1 (v1.2, Fig. 3) of *Streptomyces ambofaciens* folds into a stem-loop structure (numbered



**Fig. 3.** Secondary structure prediction for the interactions between the 16S–23S and 23S–5S spacers of the rRNA-D of *Streptomyces ambofaciens* ATCC 23877. Nucleotides are numbered starting from position 1 for each spacer region of the *Streptomyces ambofaciens* ATCC 23877 (*S. a.*) sequence. Matches between nucleotides G and U are indicated by dots. The putative double-stranded processing site (dsPS2) and the vc are boxed. The sequence differences in constant regions between *Streptomyces griseus* and *Streptomyces ambofaciens* are indicated by circled nucleotides. A, stem-loop structure in *Streptomyces coelicolor* M145 (*S. c.*). B, folding of v2.1 onto v3.1 and v2.3 onto v3.3 of *Streptomyces ambofaciens*, showing strong conservation of the stem-loop structure. C, folding of v4.2 of *Streptomyces ambofaciens*. D, nucleotide sequence alignment of the dsPS2 stems of *Streptomyces ambofaciens* (*S. a.*), *Bacillus subtilis* (*B. s.*; Ogasawara et al., 1983) and *Lactococcus lactis* (*L. l.*; Chiaruttini & Millet, 1993), showing conservation of 12 nt (grey hatched box).

I, Fig. 3). This structure is absent from that of *Streptomyces coelicolor* which presents only one stem-loop (numbered I', Fig. 3, A), the counterpart of the stem-loops I and II of *Streptomyces ambofaciens* (I and II, Fig. 3). This alteration in structure may be due to the shortened ITS in *Streptomyces coelicolor*. In contrast, *Streptomyces griseus*, which is not closely related to *Streptomyces ambofaciens* or to *Streptomyces coelicolor*, exhibits two stem-loops by the folding of the v1 (positions 14–33, Fig. 2) and the beginning of c2 (not shown on Fig. 3). As shown by Apirion & Miczak

(1993), the double-stranded processing site dsPS1 stem-loop structure involved in the 16S RNA maturation process in *Escherichia coli*, might be included in a long double-stranded region closing the 16S rRNA sequence. The dsPS1 site might then be included in *Streptomyces* within the first structures of the predicted model.

Another striking feature of the mosaic structure of the ITSs was that in *Streptomyces ambofaciens* and, to a lesser extent in *Streptomyces coelicolor* M145, v2 and v3

were present in identical combinations (v2.1/v3.1, v2.2/v3.2 and v2.3/v3.3, Fig. 3). These variable blocks bound the c3 region. The predicted secondary structure shows that the two variable blocks are involved in the formation of a stem-loop structure (numbered III on Fig. 3) exposing the c3 at the top of the structure. The c3 folds on itself to form a short stem-loop structure. This structure is also present in *Streptomyces griseus*, although v3 differed significantly (data not shown). Thus, the structural constraints may strongly select for the presence of c3 at the top of stem-loop III (Fig. 3).

Stem-loop IV is predicted to occur between c4 and vc. The formation of this stem-loop is observed regardless of the vc considered and is mostly due to extensions of single-stranded nucleotides (notably at the top of the structure).

v4 folds into a stem-loop structure (numbered V, Fig. 3). The different versions of this variable region show compensatory substitutions that allow the maintenance of the secondary structure with a stem-loop at the top of the structure (Fig. 3, C).

Predicted models are reinforced by the polymorphism observed within a given variable region (nucleotide positions double-underlined in Fig. 1) which is either located in single-stranded structures or corresponds to compensatory nucleotide substitutions (Fig. 3).

In the same way, the compensatory mutations or the G-U matches leading to the conservation of the structure of the nine different nucleotides between *Streptomyces ambofaciens*/*Streptomyces coelicolor* and *Streptomyces griseus* in c2 and c4 (Fig. 2 and circled nucleotides in Fig. 3) support the predicted models. Regions c2 (positions 78–90) and c5, constant within the *Streptomyces* species studied, fold with sequences of the 23S–5S spacer. The long stem-loop involving c5 includes a 21 nt conserved region that is thought to participate in the formation of dsPS2, which is the recognition site of RNase III required for maturation of the 23S RNA. Alignment of this region with the dsPS2 of Gram-positive bacteria *Bacillus subtilis* and *Lactococcus lactis* reveals a partial conservation of the nucleotide sequence involved in this structure (Fig. 3, D).

#### Phylogenetic use of ITSs

The high levels of variability within the 16S–23S spacer region limit its use in determining close phylogenetic relationships. Furthermore, the use of ITSs for phylogenetic tree inference is questionable, at least in the case of *Streptomyces*, due to the suggested role of gene conversion in their evolution. We propose that most nucleotide differences do not result from the accumulation of point mutations, but rather from recombination events leading to the shuffling of nucleotide blocks. Moreover, current phylogenetic tree reconstructions infer a single underlying tree topology for each informative site along a sequence. Consequently, the presence of mosaic structures may cause phylogenetic methods to produce misleading results (McGuire

*et al.*, 1997). Thus, phylogenetic distances may be overestimated, as nucleotide differences corresponding to a variable region should be considered as resulting from a single mutation event (i.e. block replacement by recombination). The same error could be made with sequences involved in lateral gene transfer. This might explain why *Streptomyces griseus* shows a high level of nucleotide divergence in its ITS D', even in constant regions (GenBank accession no. AB030571).

Constant regions of sequence are not suitable for phylogenetic studies as they include only a few informative sites. In fact, *Streptomyces ambofaciens* and *Streptomyces coelicolor* have no nucleotide substitutions in their constant regions (except for the vc), whereas nine differences were detected between the constant regions of *Streptomyces ambofaciens* and *Streptomyces griseus*. This is consistent with the phylogenetic relationships inferred from 16S rRNA sequence data (Takeuchi *et al.*, 1996).

#### Conclusions

ITSs from the *Streptomyces ambofaciens* strains determined in this study consist of five constant regions (c1–c5) and four variable regions (v1–v4); v1, v2 and v3 have three different sequences (v1.1, v1.2 and v1.3, v2.1, v2.2 and v2.3, and v3.1, v3.2 and v3.3, respectively), whereas v4 has two (v4.1 and v4.2). Combinations of variable regions lead to the existence of five different ITSs in *Streptomyces ambofaciens* DSM 40697 and in *Streptomyces ambofaciens* ATCC 23877. Eight combinations were found in total. However, no homologous pairs of rDNA loci shared the same block combination and no specific sequence exists in either of the strains. This suggests that recombination is responsible for the exchange of nucleotide blocks. Moreover, interspecific comparison shows that analogous modular structures exist in *Streptomyces coelicolor* and in *Streptomyces griseus*, with an identical succession of constant nucleotide blocks. Finally, secondary structure searches show that whatever the combination of nucleotide blocks, the same secondary structure is predicted.

#### ACKNOWLEDGEMENTS

This study was supported by the Ministère de la Recherche and by the Association pour la Recherche sur le Cancer. T. W. and V. R. were Fellows of the Ministère de la Recherche.

#### REFERENCES

- Anton, A. I., Martínez-Murcia, A. J. & Rodríguez-Valera, F. (1998). Sequence diversity in the 16S–23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *J Mol Evol* 47, 62–72.
- Anton, A. I., Martínez-Murcia, A. J. & Rodríguez-Valera, F. (1999). Intraspecific diversity of the 23S rRNA gene in the spacer region downstream in *Escherichia coli*. *J Bacteriol* 181, 2703–2709.
- Apirion, D. & Miczak, A. (1993). RNA processing in prokaryotic cells. *BioEssays* 15, 113–120.
- Berger, F., Fischer, G., Kyriacou, A., Decaris, B. & Leblond, P.

- (1996). Mapping of the ribosomal operons on the linear chromosomal DNA of *Streptomyces ambofaciens* DSM40697. *FEMS Microbiol Lett* 143, 167–173.
- Bricker, B. J. (2000). Characterization of the three ribosomal RNA operons *rrmA*, *rrmB*, and *rrmC*, from *Brucella melitensis*. *Gene* 255, 117–126.
- Chiaruttini, C. & Milet, M. (1993). Gene organization, primary structure and RNA processing analysis of a ribosomal RNA operon in *Lactococcus lactis*. *J Mol Biol* 230, 57–76.
- Christensen, H., Møller, P. L., Vogensen, F. K. & Olsen, J. E. (2000). Sequence variation of the 16S to 23S rRNA spacer region in *Salmonella enterica*. *Res Microbiol* 151, 37–42.
- Chun, J., Huq, A. & Colwell, R. R. (1999). Analysis of 16S–23S rRNA intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. *Appl Environ Microbiol* 65, 2202–2208.
- Cole, S. T. & Saint-Girons, I. (1999). Bacterial genomes – all shapes and sizes. In *Organization of the Prokaryotic Genome*, pp. 35–62. Edited by R. L. Charlebois. Washington, DC: American Society for Microbiology.
- Gürtler, V. (1999). The role of recombination and mutation in 16S–23S rDNA spacer rearrangements. *Gene* 238, 241–252.
- Gürtler, V. & Barrie, H. D. (1995). Typing of *Staphylococcus aureus* strains by PCR-amplification of variable-length 16S–23S rDNA spacer regions: characterization of spacer sequences. *Microbiology* 141, 1255–1265.
- Gürtler, V. & Mayall, B. C. (1999). rDNA spacer rearrangements and concerted evolution. *Microbiology* 145, 2–3.
- Gürtler, V. & Stanisich, V. A. (1996). New approaches to typing and identification of bacteria using the 16S–23S rDNA spacer region. *Microbiology* 142, 3–16.
- Gürtler, V., Yujun, R., Pearson, S. R., Bates, S. M. & Mayall, B. C. (1999). DNA sequence heterogeneity in the three copies of the long 16S–23S rDNA spacer of *Enterococcus faecalis* isolates. *Microbiology* 145, 1785–1796.
- Hain, T., Ward-Rainey, N., Kroppenstedt, R. M., Stackebrandt, E. & Rainey, F. A. (1997). Discrimination of *Streptomyces albidoflavus* strains based on the size and number of 16S–23S ribosomal DNA intergenic spacers. *Int J Syst Bacteriol* 47, 202–206.
- Iteman, I., Rippka, R., Tandeau de Marsac, N. & Herdman, M. (2000). Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. *Microbiology* 146, 1275–1286.
- Jumas-Bilak, E., Michaux-Charanchon, S., Bourg, G., O'Callaghan, D. & Ramuz, M. (1998). Differences in chromosome number and genome rearrangements in the genus *Brucella*. *Mol Microbiol* 27, 99–107.
- Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. & Hopwood, D. A. (2000). *Practical Streptomyces Genetics*. Norwich: The John Innes Foundation.
- La Farina, M., Stira, S., Mancuso, R. & Grisanti, C. (1996). Characterization of *Streptomyces venezuelae* ATCC 10595 rRNA gene clusters and cloning of *rrmA*. *J Bacteriol* 178, 1480–1483.
- Lan, R. & Reeves, P. R. (1998). Recombination between rRNA operons created most of the ribotype variation observed in the seventh pandemic clone of *Vibrio cholerae*. *Microbiology* 144, 1213–1221.
- Leblond, P. & Decaris, B. (1999). Unstable linear chromosomes: the case of *Streptomyces*. In *Organization of the Prokaryotic Genome*, pp. 235–261. Edited by R. L. Charlebois. Washington, DC: American Society for Microbiology.
- Leblond, P., Fischer, G., Francou, F. X., Berger, F., Guérineau, M. & Decaris, B. (1996). The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol Microbiol* 19, 261–271.
- Leblond-Bourget, N., Philippe, H., Mangin, I. & Decaris, B. (1996). 16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific *Bifidobacterium* phylogeny. *Int J Syst Bacteriol* 46, 102–111.
- Liu, S.-L. & Sanderson, K. E. (1996). Highly plastic chromosomal organization in *Salmonella typhi*. *Proc Natl Acad Sci U S A* 93, 10303–10308.
- Luz, S. P., Rodriguez-Valera, F., Lan, R. & Reeves, P. R. (1998). Variation of the ribosomal operon 16S–23S gene spacer region in representatives of *Salmonella enterica* subspecies. *J Bacteriol* 180, 2144–2151.
- Maiwald, M., Von Herbay, A., Lepp, P. W. & Relman, D. A. (2000). Organization, structure and variability of the rRNA operon of the Whipple's disease bacterium (*Tropheryma whippelii*). *J Bacteriol* 182, 3292–3297.
- Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288, 911–940.
- McGuire, G., Wright, F. & Prentice, M. J. (1997). A graphical method for detecting recombination in phylogenetic data sets. *Mol Biol Evol* 14, 1125–1131.
- Naïmi, A., Beck, G. & Branlant, C. (1997). Primary and secondary structures of rRNA spacer regions in enterococci. *Microbiology* 143, 823–834.
- Nour, M. (1998). 16S–23S and 23S–5S intergenic spacer regions of lactobacilli: nucleotide sequence, secondary structure and comparative analysis. *Res Microbiol* 149, 433–448.
- Nour, M., Naïmi, A., Beck, G. & Branlant, C. (1995). 16S–23S and 23S–5S intergenic spacer regions of *Streptococcus thermophilus* and *Streptococcus salivarius*, primary and secondary structure. *Curr Microbiol* 31, 270–278.
- Ogasawara, N., Moriya, S. & Yoshikawa, H. (1983). Structure and organization of rRNA operons in the region of the replication origin of the *Bacillus subtilis* chromosome. *Nucleic Acids Res* 11, 6301–6318.
- Pernodet, J. L., Boccard, F., Alegre, M. T., Gagnat, J. & Guérineau, M. (1989). Organization and nucleotide sequence analysis of a ribosomal RNA cluster from *Streptomyces ambofaciens*. *Gene* 79, 33–46.
- Pinnert-Sindico, S., Ninet, L., Preud'homme, J. & Cosar, C. (1955). A new antibiotic spiramycin. *Antibiot Annu* 1954–1955, 724–727.
- Privitera, A., Rappazzo, G., Sangari, P., Giannino, V., Licciardello, L. & Stefani, S. (1998). Cloning and sequencing of a 16S/23S ribosomal spacer from *Haemophilus parainfluenzae* reveals an invariant, mosaic-like organisation of sequence blocks. *FEMS Microbiol Lett* 164, 289–294.
- Redenbach, M., Kieser, H. M., Denapate, D., Eichner, A., Cullum, J. & Hopwood, D. A. (1996). A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol Microbiol* 21, 77–96.
- Rhodes, M. (1986). Genetic recombination and maps for *Streptomyces*. In *The Bacteria, a Treatise on Structure and Function*, vol. IX, *Antibiotic-Producing Streptomyces*, pp. 27–60. Edited by S. W. Queener & L. E. Day. Orlando, FL: Academic Press.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Takeuchi, T., Sawada, H., Tanaka, F. & Matsuda, I. (1996). Phylogenetic analysis of *Streptomyces* spp. causing potato scab based on 16S rRNA sequences. *Int J Syst Bacteriol* 46, 476–479.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673–4680.

---

Received 18 May 2001; revised 22 October 2001; accepted 31 October 2001.

**PUBLICATION N°2**

**End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens***

Thomas Wenner, **Virginie Roth**, Gilles Fischer, Céline Fourier, Bertrand Aigle, Bernard Decaris and Pierre Leblond

*Molecular Microbiology, sous presse*

## End-to-end fusion of linear deleted chromosomes initiates a cycle of genome instability in *Streptomyces ambofaciens*

Thomas Wenner,<sup>1†</sup> Virginie Roth,<sup>1</sup> Gilles Fischer,<sup>2</sup> Céline Fourrier,<sup>1</sup> Bertrand Aigle,<sup>1</sup> Bernard Decaris<sup>1</sup> and Pierre Leblond<sup>1\*</sup>

<sup>1</sup>Génétique et Microbiologie UMR INRA – UHP 1128, IFR110, Faculté des Sciences et Techniques, Université Henri Poincaré – Nancy 1, Boulevard des Aiguillettes BP 239, 54506 Vandœuvre-lès-Nancy – France.

<sup>2</sup>Institut Pasteur, Laboratoire de Génétique Moléculaire des Levures, 25, rue du Docteur Roux, 75724 Paris cedex 15, France.

### Summary

Two mutant strains harbouring a linear chromosome whose size reached 13 Mb (versus approximately 8 Mb for the wild type) were characterized. This chromosomal structure resulted from the fusion in inverted orientation of two chromosomes partially deleted on the same arm. The fusion occurred by illegitimate recombination between 6 bp repeats. This chromosomal structure was inherited in strict association with a high level of genetic instability (30% of mutants in a single progeny, phenomenon also called hypervariability) and chromosomal instability. In contrast, derivatives, which did not retain the chromosome fusion, showed a wild-type-like instability frequency (c. 1%). Stabilization of the chromosomal structure occurred by chromosome arm replacement or circularization. A high variability of the terminal inverted repeat (TIR) length in the rescued chromosomes (from 5 kb to approximately 1.4 Mb for linear derivatives) was observed. Mutant lineages harbouring the chromosomal fusion are characterized by a highly heterogeneous distribution of DNA in the spores, by the presence of spores without DNA as well as aberrant sporulation figures, and by the production of spores with a low germination rate. The wild-type characteristics were restored in the descendants, which lost the chromosomal fusion. Thus, the

fusion of deleted chromosomes initiates a cycle of chromosome instability sharing several levels of analogy with the behaviour of dicentric chromosomes in eukaryotes. We propose that the high instability of the fused chromosomes results from the duplication of a region involved in partitioning of the chromosomes (*parAB-oriC*).

### Introduction

*Streptomyces* are Gram-positive soil bacteria belonging to the high-GC group of the class *Actinobacteria*. They possess one of the largest bacterial chromosomes with more than 8 Mb and the largest number of genes so far discovered in a bacterium with more than 7500 protein-coding sequences as revealed by the analysis of the complete chromosome sequence of *S. coelicolor* A3(2) (Bentley *et al.*, 2002) and *S. avermitilis* (Ikeda *et al.*, 2003). The G+C content of their genome is also extreme with an average of 70% to 74% (Chater and Hopwood, 1984). In addition, the chromosomal DNA is linear and shows an invertron structure (Sakaguchi, 1990) with terminal inverted repeats (TIRs) associated with terminal proteins (Bao and Cohen, 2001). The size of the TIRs is widely variable among the linear replicons from 44 bp for the linear plasmid SLP2 of *S. lividans* (Chen *et al.*, 1993) to 550 kb for the chromosomal DNA of *S. rimosus* (Pandza *et al.*, 1997).

Genetic instability (i.e. spontaneous mutability) is widely observed among *Streptomyces* species and occurs at frequencies as high as 0.01 per spore. This instability affects numerous characters that are involved mainly in the differentiation process, expression of the secondary metabolism or response to environmental changes. In *S. ambofaciens* as well as in *S. lividans*, a two-step process generating mutants at high frequency was characterized. Hence, a dramatic increase in the frequencies of instability was observed in the progeny of mutants raised from the wild-type strain, i.e. 25% in *S. lividans* (Allenbuchner and Cullum, 1985) and 87% in *S. ambofaciens* where this phenomenon was called hypervariability (Leblond *et al.*, 1989).

Structural alterations of the chromosome are closely associated with genetic instability in the *Streptomyces*

Accepted 30 June, 2003. \*For correspondence. E-mail: leblond@nancy.inra.fr; Tel. (+33) 3 83 91 22 04; Fax (+33) 3 83 91 25 00. †Present address: Chromosome Biology Laboratory, Department of Biochemistry, National University of Galway, University Road, Galway, Ireland.



species studied to date (Altenbuchner and Cullum, 1985; Dyson and Schrempl, 1987; Leblond *et al.*, 1991). These chromosome rearrangements involve the deletion of hundreds of kilobases of the terminal regions that can include the bacterial telomere (the last tens of nucleotides forming a Y-shape secondary structure involved in terminal replication; Huang *et al.*, 1998). The rescued chromosomes are either linear or circular. In that latter case, the bulk (up to 25% of the total chromosome size in *S. ambofaciens*) of the so-called unstable region including the two telomeres is deleted and the remaining part of the chromosome is circularized (Fischer *et al.*, 1997a; Kamaoka *et al.*, 1999; Pang *et al.*, 2002; and references therein). When the chromosome remains linear, different rearrangements have been reported. First, chromosome arm replacement (i.e. terminal translocation) resulting in the duplication of large terminal areas was described in *S. ambofaciens* (Fischer *et al.*, 1998) and *S. griseus* (Uchida *et al.*, 2003). In *S. ambofaciens*, the deleted region (480 or 850 kb) was replaced by subterminal sequences originating from the intact chromosomal arm. Therefore, the TIRs varied in size from 210 kb in the wild type (Leblond *et al.*, 1996) to 450 kb and 850 kb in the two mutant strains characterized (Fischer *et al.*, 1998). Second, deletions affecting only one of the subterminal regions but not the telomere have been described (Fischer *et al.*, 1997a; Pang *et al.*, 2002). One of the consequences of these rearrangements is a decrease in the size of the TIRs, if one of the recombination events affects a sequence belonging to the TIRs. Third, DNA amplification (i.e. tandem repeats of a unit whose size can exceed 100 kb) is also frequently associated with genetic instability. In most cases well-characterized, amplification is associated with the deletion of all the sequences from the amplified locus to the end of the chromosome, the chromosomal DNA remaining linear (Pauland *et al.*, 1995; Fischer *et al.*, 1997a). In contrast, amplification associated with circularized chromosomes has been only rarely reported (Redenbach *et al.*, 2000; Catahii *et al.*, 2003).

In a plasmid model, where one telomere of the *S. rochei* plasmid pSLA2 is damaged, three kinds of rescue modes have been described (Qin and Cohen, 2002). The survivor plasmids are either circularized or remain linear by suffering non-reciprocal homologous recombination between the two terminal repeats. Alternatively, the formation of long palindromic linear plasmids that duplicate the intact telomeric sequences has been observed (Qin and Cohen, 2000).

Here we report a 'new' chromosomal rearrangement in which the loss of one chromosome arm is overcome by the fusion with another copy of the chromosome deleted on the same chromosomal arm. Thus, a typical telomere is restored at the end of a deleted chromosome. This duplicated structure (approximately 13 versus 8 Mb in the wild-type strain) enters a cycle of heritable genomic instability. The behaviour of the fusion structure is reminiscent of that of eukaryotic dicentric chromosomes (Hackett *et al.*, 2001). The inheritance of the fusion structure confers a mutator state leading to the production of genetic and phenotypic diversity within the bacterial population. In contrast, the loss of the duplicated state constitutes an exit from this instability cycle and leads to the production of chromosomes showing a high variability of TIR length.

## Results

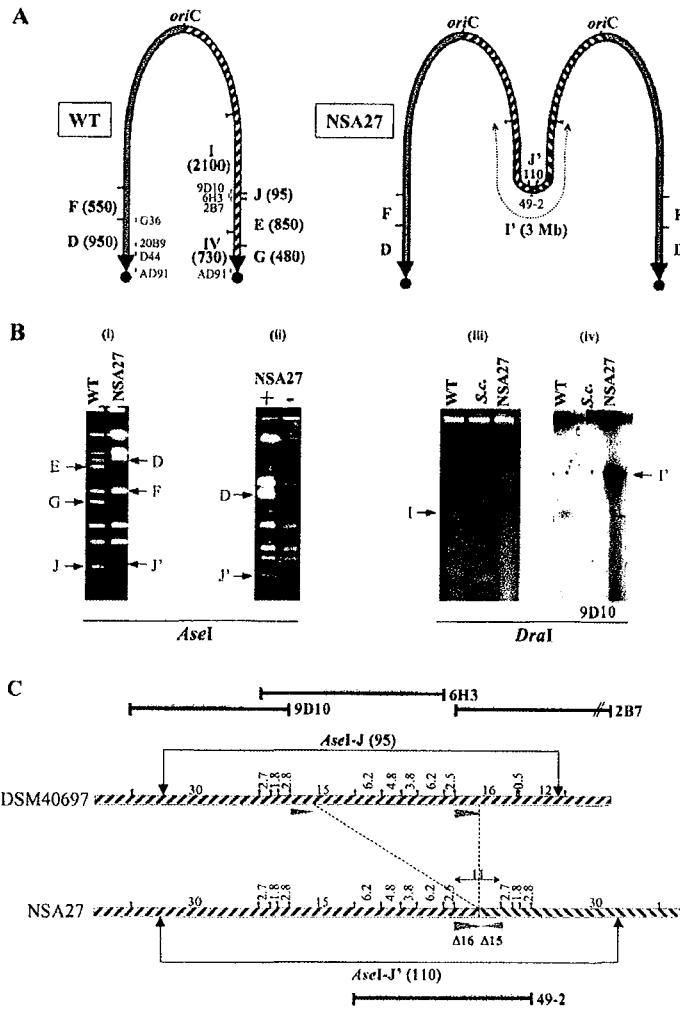
### End-to-end fusion of linear deleted chromosomes

The chromosomal structures of two spontaneous mutant strains, NSA27 and NSA65, previously isolated (Leblond *et al.*, 1991) were investigated by pulsed-field gel electrophoresis (PFGE) analysis. In NSA27, the *Asel* fragments spanning the right chromosomal arm (G-480 kb, E-850 kb and J-95 kb according to the nomenclature of Leblond *et al.*, 1996; Fig. 1A) were absent and an extra fragment called J'-110 kb was observed (Fig. 1B). From these data, the size of the deleted area was estimated to be approximately 1350 kb. When NSA27<sup>-</sup> DNA samples (prepara-

**Fig. 1.** Chromosomal structure of *Streptomyces ambofaciens* NSA27 mutant strain. A. Schematic representation of the chromosome of *S. ambofaciens* DSM40697 (wt) and NSA27. The two chromosomal arms are arbitrarily separated by the replication origin (*oriC* locus). Only the relevant parts of the *Asel* and *DraI* maps of the terminal regions are given, on the outer and inner parts of the schema respectively. Cosmids spanning the relevant regions are shown as solid bars. The black arrowheads symbolize the region corresponding to the wt TIRs. The terminal protein is symbolized by the black circle. Sizes are given in kilobases except for ' in megabase (Mb, shown by the dotted double arrows). B. PFGE analysis of *S. ambofaciens* NSA27 chromosomal DNA. (i) *Asel* pattern of *S. ambofaciens* wt and NSA27. Running conditions were 6 Vcm<sup>-1</sup> for 24 h in a 1% agarose gel using ramp pulse time from 20 to 130 s with an electrode angle of 120° (TAE 0.5x). (ii) DNA<sup>+</sup> and <sup>-</sup> protein treatment (Leblond *et al.*, 1996). (iii) *DraI* analysis of wt and NSA27 (PFGE conditions: 3 Vcm<sup>-1</sup>, 90 h, 0.6% agarose, 200–500 s, electrode angle 120°, TAE 0.5x). (iv) Hybridization of labelled DNA of cosmid 9D10 on the *DraI* patterns. PFGE size markers were *Saccharomyces cerevisiae* (*S. c.*) chromosomes (Bio-Rad). C. Interpretative representation of the fusion region in mutant strain NSA27. Representative cosmids are indicated as bold lines. The partial *Bam*HI (*B*) restriction maps of the regions involved in the fusion are deduced from overlapping cosmids retrieved from wt and NSA27 genomic libraries. Sizes of fragments are shown in kb. The dotted lines indicate the recombination points. The filled arrowheads symbolize the partially deleted *Bam*HI fragments involved in the formation of the fusion. The direction of the hatching symbolizes the relative orientation of the fused fragments. Cosmid 2B7 is shown truncated (*t*).

tion without proteinase treatment) were electrophoresed, only the migration of the *Asel* fragment D-850 kb corresponding to the intact left chromosomal arm was impaired (Fig. 1B). Thus, the deletion of the right chromosomal arm left the NSA27 chromosome linear. In order to determine the structure of the new *Asel*-J' fragment, a cosmid library was constructed from the total genomic DNA of NSA27 and screened using the purified DNA of the J'-110 kb used as a probe. Only cosmids harbouring sequences

originating from the *Asel* J-95 kb fragment were retrieved. The *Bam*I restriction map of the J'-110 kb could be deduced from overlapping cosmids and indicated that this fragment resulted from the fusion in opposite orientation of two partially deleted J-95 kb fragments (Fig. 1C). The two J-95 kb fragments are deleted to different extents, approximately 25 and 55 kb. The new 11 kb *Bam*I fragment (cloned in cosmid 49-2) consisted of the fusion of the two partially deleted 16 and 15 kb *Bam*I fragments



(Fig. 1C). Together, these data suggested that the NSA27 chromosomal structure resulted from the fusion in opposite orientation of two chromosomes partially deleted on the same arm (Fig. 1A). To test this hypothesis, the structure of the NSA27 chromosome was further investigated using *DraI* (the second enzyme used to map the *S. ambotaciens* chromosome, Leblond *et al.*, 1996). The two terminal fragments IV (730 kb) and I (2100 kb) were affected (Fig. 1C, fragment IV is not visible using this PFGE resolution) that is consistent with the deletion affecting the right chromosomal arm over 1.3 Mb. A new *DraI* fragment, named I', whose size was estimated to be 3 Mb was shown to derive from the 2.1 Mb by hybridization with cosmid 9D10 DNA (Fig. 1). This fragment is interpreted as the result of the fusion in opposite orientation of two partially deleted *DraI*-I (2.1 Mb fragments) according to the structure deduced from the analysis of *Asel*-J' (Fig. 1C). Furthermore, the fusion hypothesis implies that all the restriction fragments except for the one including the junction should exhibit a double relative intensity. Both *Asel*-J' and *DraI*-I' fragments showed lower intensities than expected. However, for *DraI*-I', these data might be inconclusive because of the statistical degradation affecting preferentially large DNA fragments in PFGE experiments. This underrepresentation was even more marked in some DNA samples (by a factor 2–4), presumably because of the strong chromosomal instability affecting the fusion strains (see below).

Similar data were obtained for the NSA65 strain (Fig. 2). This deletion affected the right chromosomal arm over approximately 1330 kb. The *Asel* fragments G-480 kb and E-850 kb were missing but no extra fragment was detected (Fig. 2B). The deletion end-point was mapped in a 12 kb *Bam*HI fragment and a new 18.5 kb fragment was detected (Fig. 2C). The fragment consisted of the fusion of two partially deleted 12 kb *Bam*HI fragments as revealed by restriction analysis and hybridization with specific probes. This structure would generate a J'-94.5 kb fusion that might remain unresolved from the intact J-95 kb under the PFGE conditions applied (Fig. 2B). The presence of a double band J-J' was further suggested by densitometric analysis revealing an equivalent relative intensity between fragments *Asel*-I (210 kb) and 95 kb (not shown). In addition, the *DraI* pattern, like NSA27, revealed a new fragment (I') of about 3 Mb consistent with the duplication of the remaining part of the *DraI*-I fragment (2100 kb, Fig. 2B).

Thus, the sizes of the NSA27 and NSA65 mutant chromosomes are believed to exceed 13 versus 8 Mb in the wild-type strain with duplicated regions of about 6500 kb constituting the TIRs.

#### *Illegitimate recombination involved in the NSA27 fusion*

The NSA27 fusion point was determined by sequencing

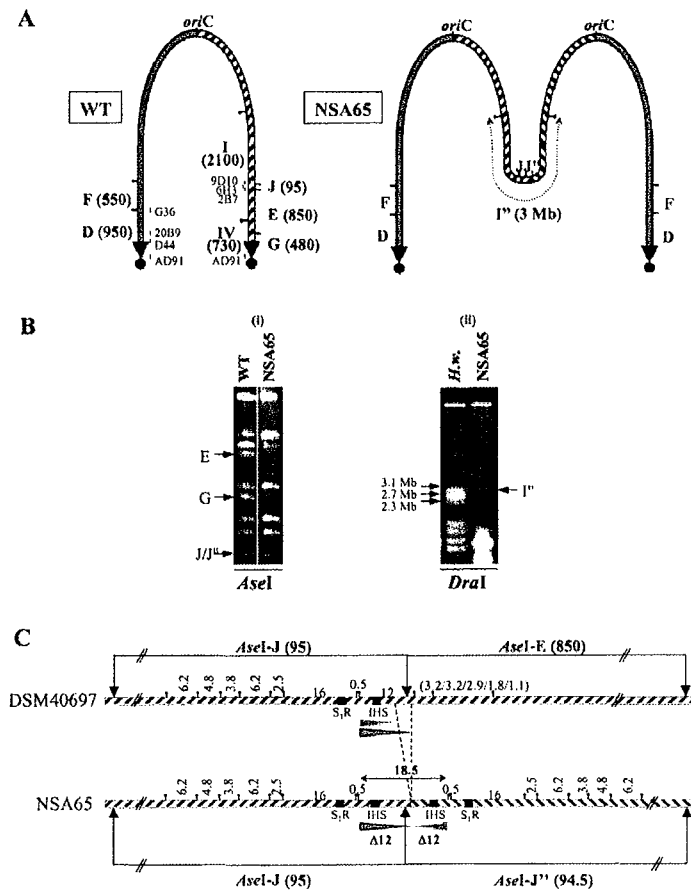
a 1.6 kb *Pst*I-*Sph*I fragment mapped in the 11 kb *Bam*HI fragment including the junction (Fig. 3A). A strong homology, 95% of identity over the first 856 nt, was detected with a part of an open reading frame (ORF) encoding a putative ABC transporter of *S. coelicolor* A3(2) (Huang *et al.*, 2001). The homology stopped abruptly beyond this point suggesting that the recombination point was localized in the vicinity and that the recombination event disrupted the corresponding ORF. Two oligonucleotides (P<sub>2B7</sub> and P<sub>6H3</sub>) were designed on both sides of the putative recombination point and the wild-type sequences involved in the recombination event were determined from the cosmids 2B7 and 6H3 (Fig. 3A). The sequence deduced from cosmid 2B7 (269 nt) confirmed the homology with the previously identified transporter. No significant homology could be found for the whole sequence obtained from cosmid 6H3 (503 nt). The junction sequence was compared with the right and left wild-type sequences and revealed that an illegitimate recombination event had occurred within a 6 bp microhomology (Fig. 3B). The two sequences mapped in direct orientation within J-95 kb as deduced from the combination of the restriction map of the wild-type sequences and that of the fusion fragment, while they are fused in opposite orientations in the NSA27 chromosome.

#### *Genomic instability of the chromosomal fusions*

We investigated the inheritance of the fused chromosome structure and analysed samples of clones randomly selected in both NSA27, NSA27r (a marked NSA27 strain, see *Experimental procedures*) and NSA65 progeny (Table 1). While the progeny of the wild-type strain inherited the parental wild-type chromosomal structure (24 clones tested), rearrangement of the parental structure was highly frequent in both NSA27 and NSA65 progenies (non-parental chromosome or NP). Thus, 11 out of 19 and 7 out of 15 colonies tested for NSA27 and NSA65, respectively, showed a NP structure. For NSA27r, the ratio of rearranged chromosomes was even higher (15 of 17 tested). This might be due to the selection procedure implying several cycles of sporulation before the analysis of the chromosomal structure of the descendants. The NP chromosomal structure can consist of either linear (L) or circular (C) chromosomes. Linear chromosomes can correspond to the fusion of two partially deleted chromosomes, however, with a different *Asel* junction (e.g. J'-110 kb in NSA27). These descents are called NF for new fusion. Chromosomal linearity can also be reached with the loss of the fusion state (LF for loss of fusion). In those cases, a wide diversity of TIR length was observed.

#### *Characterization of chromosomal circularization*

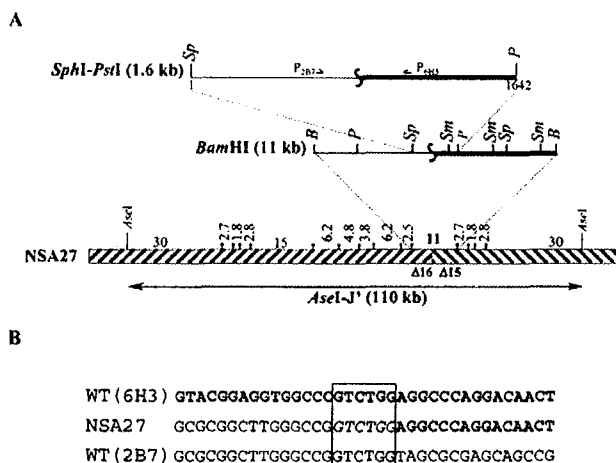
The circularized chromosomes have been studied in sev-



**Fig. 2.** Chromosomal structure of *S. ambotaciens* NSA65 mutant strain. **A.** Schematic representation of the chromosome of *S. ambotaciens* wt and NSA65. Symbols are as in Fig. 1A. Sizes are given in kilobases except for I'' (shown by the dotted double arrows). **B.** PFGE analysis of *S. ambotaciens* NSA65 chromosomal DNA. (i) *AseI* pattern of *S. ambotaciens* wt and NSA65 (PFGE conditions: 6 Vcm<sup>-1</sup>, 24 h, 0.9% agarose, 40–160 s, electrode angle 120°, TAE 0.5x). (ii) *DraI* analysis of wt and NSA65 (PFGE conditions: 3 Vcm<sup>-1</sup>, 48 h, 0.6% agarose, 200–500 s, electrode angle 94°, TAE 0.5x). PFGE size marker was *Hansenula wingei* (*H. w.*) size ladder (Bio-Rad). **C.** Interpretative representation of the fusion region in mutant strains NSA65. Symbols used are as in Fig. 1C. When the relative order of fragments was not defined, their sizes (in kb) are indicated into brackets. Motifs IHS and S<sub>1</sub>R were characterized by Aubert *et al.* (1993) and revealed the duplication in the 18.5 kb *Bam*HI fragment of strain NSA65. The *AseI*-J, -J' and -E are shown partially (// indicates that the map is truncated).

eral *Streptomyces* species and will not be analysed further. For 5, 2 and 3 derivatives of NSA27, NSA27r and NSA65, respectively, chromosome circularization was deduced (Table 1). Circularization is inferred from three simultaneous criteria as previously reported (Fischer *et al.*, 1997a). First, the residual terminal *AseI* fragment (*AseI* D-950 kb, Fig. 1A) should be missing. Second, a new extra

*AseI* fragment showing a relative intensity consistent with its presence in stoichiometry in the chromosome should appear. This fragment is interpreted as the junction after circularization. Third, no hybridization with the cosmid AD91 (Fig. 1A) including the proximal parts of the TIRs (with the exception of the non-clonable terminal *Bam*HI fragment, that is about 2.7 kb) should be observed.



**Fig. 3.** Analysis of the recombination point in *S. amblofaciens* NSA27. **A.** Restriction maps of the 11 kb *BamHI* fragment including the recombination point (**B**: *BamHI*; *P*: *PstI*; *Sp*: *SphI*; *Sm*: *SmaI*). This fragment consists of the fusion of the partially deleted 16 and 15 kb *BamHI* fragments from cosmids 2B7 (thin line) and 6H3 (bold line) respectively. The oligonucleotides (*P*<sub>6H3</sub> and *P*<sub>2B7</sub>) used as primers in PCR and sequencing experiments are symbolized by half arrows (see *Experimental procedures*). The sequence of the *SphI-PstI* fragment was fully determined as 1642 nucleotides (GenBank Accession Number AY227908). *P*<sub>6H3</sub> and *P*<sub>2B7</sub> primers raised sequences from cosmids 6H3 and 2B7 of 503 nucleotides (AY227910) and 269 nucleotides (AY227909) respectively. **B.** Sequences involved in the rearrangement in NSA27 and wild type. The 6 bp sequence involved in the illegitimate recombination event is boxed. Bold and thin fonts correspond to nucleotide sequences raised from cosmids 6H3 and 2B7 respectively.

**Characterization of new chromosomal fusions**

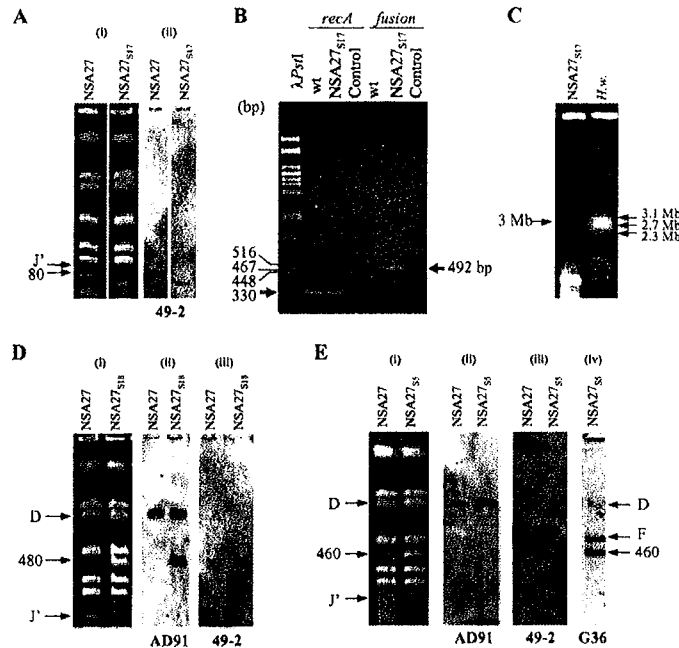
A new *AseI* fusion fragment was characterized in 1 and 3 derivatives of NSA27 and NSA65 respectively (Table 1). No such derivative was found in the progeny of NSA27r. As an example, in strain NSA27<sub>S17</sub>, the J'-110 kb fragment is missing while a new fragment of about 80 kb is detected (Fig. 4A). This fragment appeared to be underrepresented taking its relative intensity into account as for J'-110 kb in NSA27. Further, labelled DNA of cosmid 49-2 (overlap-

ping the fusion in NSA27, Fig. 1A and C) hybridized with this new fragment (Fig. 4A). The presence of the NSA27 recombination point was further confirmed by the observation of a PCR product of the expected size (492 bp) using *P*<sub>2B7</sub> and *P*<sub>6H3</sub> as primers (Figs 3 and 4B). Finally, *DraI* analysis confirmed the fused structure by the detection of a fragment of about 3 Mb (Fig. 4C). These data showed that NSA27<sub>S17</sub> (i) is a true derivative of NSA27, (ii) is rearranged compared with the parental strain and (iii) retains the fusion structure (Fig. 5).

**Table 1.** Genomic and phenotypic characterization of the wild type, NSA27, NSA27r, NSA65 and their derivatives.

Strain	Estimated TIR size (kb)	n	P	C	NP			Genetic instability	Germination rate
					LF	NF	L		
DSM40697	210	24	24	0	0	0	1.4 ± 1.1%	66%	
NSA27	6500	19	8	5	5	1	30.2 ± 8.1%	9%	
NSA27r	6500	17	2	2	13	0	ND	ND	
NSA65	6500	15	8	3	1	3	36.4 ± 8.1%	25.6%	
NSA27 <sub>S10</sub> P	6500	10	9	1	0	0	27 ± 6.3%	6%	
NSA65 <sub>S5</sub>	6500	10	9	1	0	0	26 ± 6.5%	16%	
NSA27 <sub>S17</sub> NF	6500	20	9	4	7	0	22.5 ± 5.9%	5%	
NSA65 <sub>S4</sub>	6500	9	3	4	2	0	30.8 ± 5.8%	8%	
NSA65 <sub>S20</sub>	6500	10	8	2	0	0	26.9 ± 7.0%	16%	
NSA27 <sub>S5</sub> LF	1300	10	10	0	0	0	0.45 ± 0.9%	77%	
NSA27 <sub>S18</sub>	370	10	10	0	0	0	0.7 ± 1.0%	71%	
NSA27 <sub>S13</sub>	80	10	10	0	0	0	1.1% ± 1.2%	40%	
NSA27 <sub>S11</sub>	85	10	10	0	0	0	2.1% ± 1.8%	45%	
NSA65 <sub>S1853</sub>	1005	10	10	0	0	0	0.18 ± 0.4%	76%	

The strains are sorted according to their deduced chromosomal structure: parental (P), new fusion (NF) which means conservation of the fusion structure with a new *AseI* junction fragment and loss of fusion (LF) which means conservation of a linear chromosome, however, with the loss of the parental chromosomal fusion. The estimated size of the TIRs is indicated in kilobases and is deduced from PFGE and hybridization experiments (see text). NP: non-parental chromosome structure; C: circular. Linear chromosomes (L) are either NF or LF. The percentages of genetic instability correspond to the frequency of colonies with a non-parental phenotype. The total number of colonies analysed ranged from 140 to 487. The germination rates are indicated after 4 days of growth.



**Fig. 4.** Chromosomal structures of derivatives of *S. amblofaciens* NSA27. **A.** *S. amblofaciens* NSA27<sub>S17</sub>. (i) *Ase*I pattern of the chromosome of NSA27 and NSA27<sub>S17</sub> (PFGE conditions: identical to Fig. 1A) and (ii) hybridization with labelled DNA of cosmid 49-2 (overlapping the recombination point in NSA27, Fig. 1A). **B.** PCR amplification of the fusion in NSA27<sub>S17</sub> (492 bp). Amplification of the *recA* gene was used as a positive control (330 bp, see *Experimental procedures*). Running conditions: 5 Vcm<sup>-1</sup>, 2 h, 2% agarose, TAE 1x. The negative control consists in a PCR reaction on water. **C.** *Dra*I analysis of NSA27<sub>S17</sub> (PFGE conditions: identical to Fig. 1B). Size markers are the same than in Fig. 2. **D.** *S. amblofaciens* NSA27<sub>S18</sub>. (i) *Ase*I pattern of the chromosome of NSA27 and NSA27<sub>S18</sub> (PFGE conditions: identical to Fig. 1A) and hybridization with labelled DNA of cosmids AD91 (ii) and 49-2 (iii). **E.** *S. amblofaciens* NSA27<sub>S5</sub>. (i) *Ase*I pattern of the chromosome of NSA27 and NSA27<sub>S5</sub> (PFGE conditions: identical to Fig. 1A) and hybridization with labelled DNA of cosmids AD91 (ii), 49-2 (iii), and G36 (iv) (see localizations in Fig. 1).

**Loss of chromosomal fusion**

The loss of chromosomal fusion was characterized in 5 and 1 derivatives of NSA27 and NSA65 respectively (Table 1). This type of chromosomal structure was significantly overrepresented in the progeny of NSA27r (13 clones of 17 tested) compared with both NSA27 and NSA65 (*P* < 0.05). As an illustration, the *Ase*I pattern of strain NSA27<sub>S18</sub> is characterized by the loss of the J' fragment and by the presence of a new 480 kb fragment (Fig. 4D). This fragment showed a relative intensity equivalent to the other fragments. It hybridized with AD91 including sequences belonging to the TIRs, with 49-2 including the fusion (Fig. 4D) and with D44 (not shown) which maps on the left chromosomal arm but out of the TIRs (Fig. 1A). The 480 kb fragment also showed a retardation of its PFGE

migration characteristic of the association with terminal proteins (not shown). Thus, the loss of the chromosome fusion was accompanied by conservation of the linear structure. The new junction consisted of a 5 kb *Bam*HI fragment that hybridized with both cosmids 20B9 (belonging to *Ase*I-D, Fig. 1A) and 6H3 (*Ase*I-J) DNAs. Thus in NSA27<sub>S18</sub>, about 370 kb of DNA remained from the parental chromosomal fusion leading to new TIRs of approximately 370 kb in size in that mutant strain (Fig. 5).

Another example of the chromosomal fusion loss was characterized in strain NSA27<sub>S5</sub>. In this mutant, a new fragment of about 460 kb hybridizing with cosmid 49-2 but not with AD 91 was detected (Fig. 4E). This fragment showed a relative intensity consistent with its presence in stoichiometry in the chromosome. Furthermore, no PFGE retardation was noted for the new 460 kb in DNA<sup>-1</sup> sam-

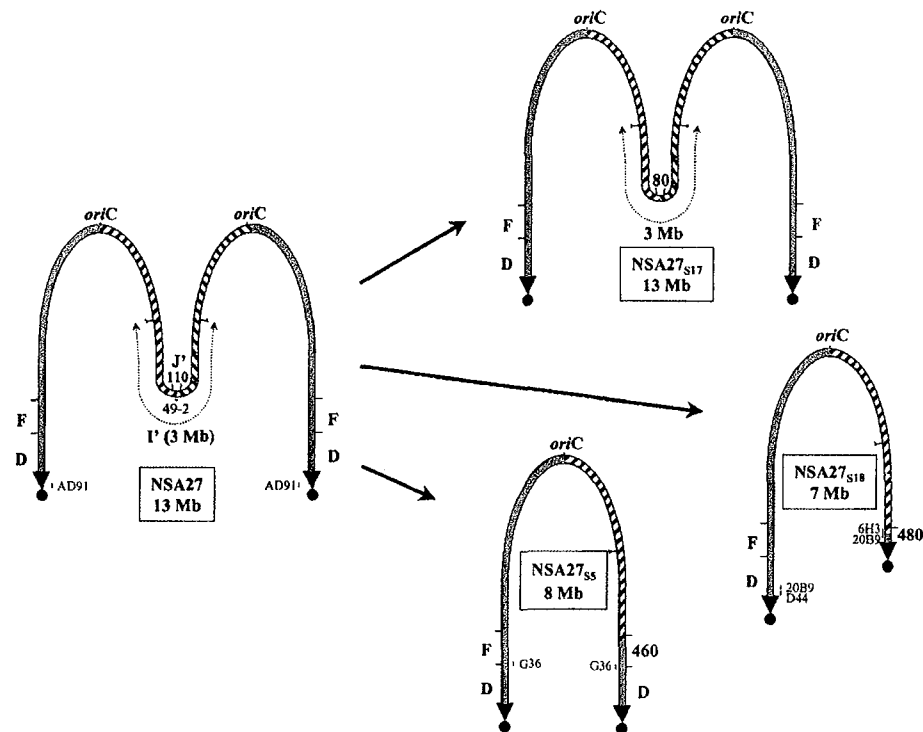


Fig. 5. Interpretative representation of the chromosomal structures of NSA27 and of its derivatives. Symbols are identical to Figs 1 and 2. Sizes are given in kilobases unless otherwise indicated.

ple compared to that of DNA<sup>+</sup> (not shown). In contrast, the *AseI*-D fragment (950 kb) showed an intensity higher than in the parental strain NSA27 (as shown by densitometric analysis and confirmed by hybridization with AD91 DNA) consistent with its being present in two copies. Hybridization with cosmid G36 (linking clone between *AseI*-F and -D fragments, Fig. 1A) revealed, in addition to D and F, the new 460 kb fragment (Fig. 4E). Thus, in that strain, the loss of the parental chromosomal fusion leaves about 1300 kb of duplicated DNA (Fig. 5).

The estimated size of the duplicated areas on the chromosome of the LF progeny of NSA27, NSA27<sub>r</sub> and NSA65 varied from 5 to 1430 kb for 28 descendants analysed in total.

#### A cycle of chromosome instability

The progeny of 6 and 4 derivatives of NSA27 and NSA65, respectively, all harbouring a linear chromo-

some, were investigated for genomic instability (Table 1). The inheritance of P (parental), NF (new fusion) and LF (loss of fusion) chromosomal structures was examined. The two first classes (P and NF) grouped together strains harbouring a chromosomal fusion and all their derivatives exhibited chromosome instability (P: NSA27<sub>S10</sub>, NSA65<sub>S5</sub>; NF: NSA27<sub>S17</sub>, NSA65<sub>S4</sub> and NSA65<sub>S20</sub>). Thus, one of 10 clones tested for NSA27<sub>S10</sub> and NSA65<sub>S5</sub> harboured a NP chromosome. The progeny of NSA27<sub>S17</sub> showed 11 NP of 20 clones tested. For NSA65<sub>S4</sub>, 6 out of 9 clones were rearranged. In the same way, 2 out of 10 NSA65<sub>S20</sub> derivatives showed a NP chromosome (Table 1). This strong genomic instability contrasted with the inheritance of the parental chromosome structure in the third category of mutant strains (LF). Thus, for the derivatives of all five LF (NSA27<sub>S5</sub>, NSA27<sub>S18</sub>, NSA27<sub>S10</sub>, NSA27<sub>S11</sub> and NSA65<sub>S19S3</sub>), no NP chromosomes were observed among 10 colonies tested for each strain.

These data strongly distinguished chromosome structures relative to their stability. Thus, the fusion state is strictly associated with genomic instability. In contrast, derivatives that have lost the chromosomal fusion recovered an inheritance pattern indistinguishable from the wild-type strain. This correlation was established in two independent genetic backgrounds (i.e. NSA27 or NSA65). Further, genomic instability cannot be ascribed to the large deletion associated with the fusion as the loss of the fusion conferred a relative 'chromosomal stabilization' despite retaining the parental terminal deletion.

#### *Phenotypic characterization of the chromosomal fusion progeny*

**Genetic instability and hypervariability.** In our experiment, the wild-type strain raised about 1.4% of phenotypic variants which is in good agreement with previous studies (Leblond *et al.*, 1991; Fischer *et al.*, 1997b). In the course of the same experiment, the NSA27 and NSA65 strains, respectively, raised 30.2% and 36.4% of variants (Table 1). In addition, a strict correlation was drawn between the chromosomal status and the level of genetic instability in the progeny of the two parental strains. Thus, on one hand, P and NF strains showed very high instability levels (ranging from 22.5% to 30.8%). On the other hand, LF strains showed instability frequencies ranging from 0.18% to 2.1% (Table 1). These frequencies are of the same order as the basic genetic instability level pointed in the wild-type strain.

In addition, the high genetic instability levels are associated with the production of several types of variants, a phenomenon reminiscent of what was called hypervariability in early studies of genetic instability in *S. ambotaciens* (Leblond *et al.*, 1989). Hypervariability was shown to be a heritable character in some descendants while some recovered a relative stability. Here, hypervariability is strictly associated with the inheritance of a chromosomal fusion while 'stabilization' is associated with its loss.

#### *Correlation between chromosomal fusion and sporulation impairment*

**DAPI-staining of spore chains.** DNA content of spores was estimated by measuring the intensity of the fluorescence after DAPI-staining (Kieser *et al.*, 2000). For that purpose, chains of spores were selected by phase-contrast microscopy in a single focal plane. Then, the intensity of fluorescence of each spore was collected, grouped into classes (see *Experimental procedures*) and the standard deviation calculated within the spore sampling (Fig. 6A). All but one distribution (i.e. NSA65) exhibited a normal distribution ( $P < 0.01$ ) allowing the comparison of their

standard deviation. Then, strains NSA27 and NSA27<sub>S17</sub> showed a significantly wider distribution than the wild-type strain and other derivatives revealing a strong imbalance of the DNA content in spores. For NSA65, the non-Gaussian distribution ( $P < 0.01$ ) revealed the production of distinct types of spores (e.g. spores devoid of DNA) and did not allow further comparison of the sampling. In contrast, derivatives that have lost the chromosomal fusion (NSA27<sub>SS</sub>, NSA27<sub>S18</sub> and NSA65<sub>S19SS</sub>) showed a fluorescent intensity distribution narrower than their parental strains, although wider than the wild type ( $P < 0.01$ ). In addition, they showed a significantly different distribution between them ( $P < 0.01$ ). All these data confirmed the direct observations of the photographs revealing the heterogeneous distribution of the DNA content within the spore chains of the mutants harbouring a chromosomal fusion while the loss of the fusion was associated with the recovery of a homogeneous repartition (Fig. 6B).

Furthermore, transmission electron microscopy of NSA27 and NSA65 spore chains revealed the presence of 'ghost' spores as well as a strong heterogeneity in size (Fig. 7B). 1/66 and 19/108 spores were devoid of DNA in NSA27 and NSA65 respectively. In addition, frequent aberrant sporulation figures (Fig. 7D) were observed only in these mutant strains (7 and 13 in 66 and 108 spores for NSA27 and NSA65 respectively). A cellular component, which might correspond to DNA, was frequently observed in the septation zone and seemed to prevent the correct separation of adjacent spores (Fig. 7D). In contrast, neither such figures nor ghost spores were observed in the wild-type strain or in the mutant derivatives that had lost the chromosomal fusion (among 46, 20 and 49 spores observed for the wild type, NSA27<sub>S18</sub> and NSA27<sub>S19SS</sub> strain respectively). An example of spore chain in NSA27<sub>S19SS</sub> is shown in Fig. 7C. Thus, the apparent impairment of the sporulation process is strictly associated ( $P < 0.05$ ) with the chromosomal fusion as it disappeared concomitantly with its loss. In addition, the size of the TIRs (from 210 kb in the wild type to 1005 kb for NSA65<sub>S19SS</sub>) does not seem to influence significantly the sporulation process.

#### *Germination rates*

The impairment of the sporulation process was further supported by the determination of the germination rates (Table 1, see *Experimental procedures*). These ratios were determined several times and varied widely as a result of experimental limitations (the values of one test are given in Table 1). However, taking into account the sets of values, a clear tendency appeared: while strains harbouring a chromosomal fusion exhibited low germination ratios, strains that have lost the fusion state seemed to have recovered wild type rates. The low germination



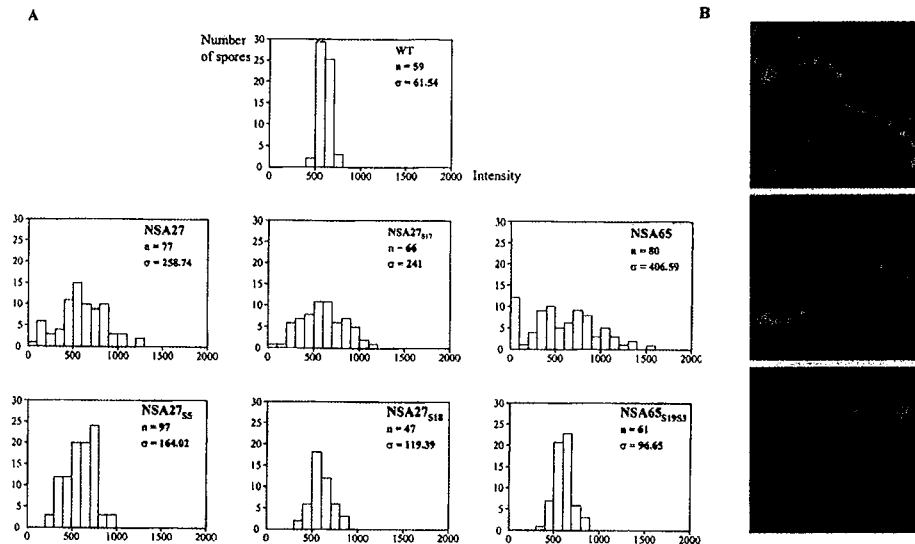


Fig. 6. DAPI-staining analysis of spores.

A. Distribution of the intensities in samples of spores of the wt (upper), mutant strains harbouring a chromosomal fusion (middle) and derivatives which have lost the parental fusion (bottom). The size of the sample is  $n$ . The standard deviation is  $\sigma$ . All the distributions have been centred on an arbitrarily chosen value of intensity (i.e. 600).

B. Typical DAPI-staining of spore chains of the wt (upper), NSA65 (middle) and NSA65S19S3 (bottom) strains (see *Experimental procedures*). Magnification was 1000 times.

rates of the fusion mutants are expected to result mainly from the highly variable DNA content of the spores. The frequencies of ghost spores cannot alone explain the high proportion of non-germinating spores.

### Discussion

In this work, we characterized a new DNA rearrangement affecting the terminal regions of the chromosome and associated with genetic instability in *Streptomyces*. This rearrangement corresponded to the formation of a large terminal deletion accompanied by the duplication of almost all the remaining part of the chromosome. This structure includes a 30 kb unique region flanked by extremely long inverted repeats of about 6.5 Mb in NSA27 (Fig. 1C). In NSA65, the unique region is even shorter with less than 0.5 kb (Fig. 2C). The consequence of this chromosomal end-to-end fusion is the production of mutant chromosomes exceeding 13 Mb of DNA versus 8 Mb in the wild-type strain. This new type of replacement of chromosomal arm involves illegitimate recombination while previous reports implied homologous recombination occurring between duplicated

genes (Fischer *et al.*, 1998; Uchida *et al.*, 2003). Qin and Cohen (2000) previously reported the formation of long-palindromic linear plasmids. However, this rearrangement results from intra-strand annealing followed by strand extension, that is a non-recombinational mechanism.

The molecular analysis of the progeny of *S. ambotaciens* NSA27 and NSA65 mutant strains allows us to propose a new global hypothesis of genomic instability where the fusion of partially deleted chromosomes plays a central part. Indeed, the analysis of the chromosomal diversity of their progeny allowed us to observe all the DNA rearrangements described so far in *S. ambotaciens* chromosomal instability studies. The only rearrangement that was not observed in that study is DNA amplification, which was previously correlated with hypervariability in *S. ambotaciens* (Leblond *et al.*, 1989). However, the two mutant strains NSA27 and NSA65 are either partially or totally deleted for the two main amplifiable loci in *S. ambotaciens* characterized and thus might be unable to suffer DNA amplification frequently.

The end-to-end fusion of two deleted chromosomes generates a high level of chromosomal instability

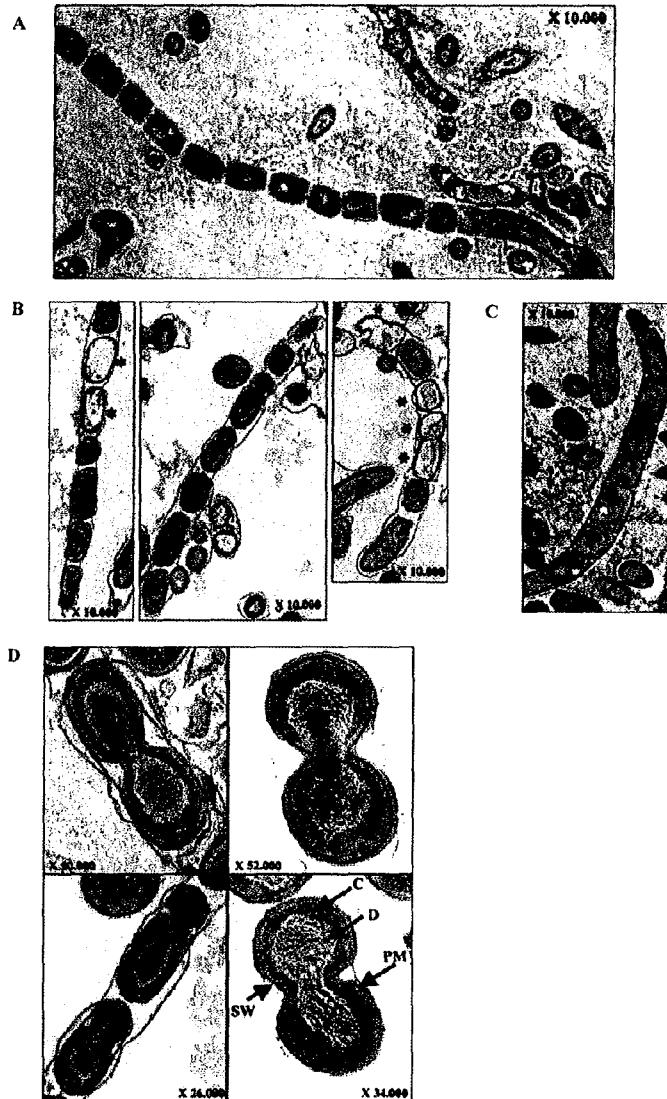


Fig. 7. Transmission electronic micrographs of chains of spores of the wt (A), NSA65 (B) and NSA65<sub>S1953</sub> (C) strains. An asterisk indicates a ghost spore. Close-up shots of examples of impaired sporulation events are shown in (D) (NSA65 and NSA27 left and right pictures respectively). Magnifications are indicated on each shot. D: DNA; C: cytoplasm, PM: plasmic membrane, SW: spore wall.

correlated with genetic instability. Indeed, the frequencies of variant phenotypes in the progeny of NSA27 and NSA65 were estimated to be 36.4% and 30.2% respectively. This is reminiscent of a specific instability phenomenon previously described in *S. ambotaciens*, which was called hypervariability (Leblond *et al.*, 1989). Hypervariability is a high-level instability level observed in the progeny of mutants raised from the basic instability (in *S. ambotaciens*, from pigment-defective colonies). It is defined as the appearance of several mutant phenotypes in the progeny of a single colony. A similar two-step mechanism was reported in *S. lividans* where mutants sensitive to chloramphenicol appear from the wild type with a frequency of 0.5% and give rise to arginine auxotrophs at a frequency of approximately 25% (Altenbuchner and Cullum, 1984). In the current study, we found that hypervariability is strictly and concomitantly inherited with the chromosomal fusion state. Indeed, the parental strains NSA27 and NSA65 and all their descents that have retained the fusion state (parental or rearranged) perpetuate the hypervariability. The loss of the fusion is associated with the recovery of a wild-type-like genetic instability level. We conclude that chromosomal end-to-end fusion initiates a cycle of genomic and genetic instability in *S. ambotaciens*. The high genomic instability is associated with the duplication of a major part of the chromosome although large terminal repeats do not trigger chromosomal instability at comparable frequency. The exit of the cycle of genomic instability occurs by the loss of the fusion state and either circularization or recovery of the linear structure. Circular chromosomes correspond to intermediate levels of genetic and genomic instability as reported earlier (Fischer *et al.*, 1997b).

The chromosomal instability cycle described here leads to a high variability of the terminal regions. The high frequency of DNA rearrangements in the terminal regions might also reveal a recombinogenic tendency and favour the integration of new genetic information. Then, this cycle might constitute a mechanism of rapid evolution of the chromosomal terminal regions (Leblond and Decaris, 1999) including up to 25% of the approximately 8000 ORFs of the chromosome.

It is tempting to relate the cycle of chromosome instability to the presence of a duplicated locus involved in the chromosome partitioning during sporulation. This locus, in single copy in the wild type as well as in the derivatives that have lost the chromosomal fusion, would be involved in the correct distribution of sister chromatids in individual spores. This process was the best characterized during *Bacillus subtilis* endospore formation (Lewis, 2001). Thus, when the chromosomal replication is initiated, a protein complex (SpoJ/Soj) binds to each sister chromatid at specific sites scattered in an area covering approximately 20% of the chromosome including the replication origin

(Wheeler and Shapiro, 1997). This DNA/protein complex is then pulled towards the poles of the cell by a mechanism, which remains to be elucidated. In *S. coelicolor* A3(2), counterparts of the genes *spoJ* and *soj*, named *parB* and *parA*, respectively, map near the origin of replication (Kim *et al.*, 2000). The interruption of the *parB* gene led to the appearance of spores with little or no DNA content, suggesting a role of the ParB protein in chromosome partitioning in *Streptomyces*. The chromosome of *S. coelicolor* A3(2) contains 24 fixation sites for ParB among which 22 are included within 100 kb flanking the replication origin (Jakimowicz *et al.*, 2002). Thus, ParA and ParB might play a central role in chromosome partitioning in *Streptomyces*. In *S. ambotaciens* NSA27, NSA65 and their fusion derivatives, two partitioning regions would be present on a single chromosome and might suffer physical friction towards the poles of the cell in division (sporulation). In contrast, derivatives that have lost the fusion would retain a single partitioning function and recover a correct segregation process.

In addition, the presence of two replication origins on the fused chromosomes might generate a strong genomic instability at the meeting point of the two forks, which is the fusion region. The genomic instability associated with pauses of the replication forks is well documented in *Escherichia coli* (Bidnenko *et al.*, 2002) and might be involved in the high plasticity of the terminus of replication. Hence, chromosome instability might occur during vegetative (intense replication phase) as well as during the sporulation process.

The cycle of genomic instability uncovered in *S. ambotaciens* associated with the end-to-end fusion of deleted chromosomes shares several features with the behaviour and consequences of dicentric chromosomes in eukaryotes, the dicentric chromosomes being involved in the so-called break-fusion-bridge cycle (BFB) as initially described by McClintock (1951). Thus, the initiation step of both cycles is the formation of a chromosomal fusion. In eukaryotes, these fusions result from the formation of double-strand breaks (originally involving transposition events) followed by non-homologous end joining (Moore and Haber, 1996). The formation of a dicentric chromosome represents the beginning of a cycle where the moving of the homologous centromeres towards the opposite poles of an achromatic spindle leads to the breakage of the DNA bridge. These newly formed DNA extremities will recombine further and give rise to a new dicentric chromosome. When the two centromeres are pulled towards the same pole, the dicentric chromosome is inherited. Exit from the cycle requires the *de novo* addition of telomeric sequences or non-reciprocal translocations. Obviously, a significant cell death ratio is associated with such phenomena. In *S. ambotaciens*, the appearance of new fusions in the progeny of NSA27 and NSA65 is reminis-

cent of the breakage-fusion step of the BFB eukaryote cycle. The generation of a wide range of terminal inverted repeats as well as circularization can be interpreted as salvage mechanisms (double-strand break repair mechanisms) and exits from the cycle. At the phenotypic level, very low germination ratios characterize the fusion strains and might result from impairment of segregation of the duplicated chromosomes. This hypothesis is supported by the observation of abnormal figures of sporulation in the only fused strains (Fig. 7). These figures might result from an inability to achieve complete spore differentiation when the chromosomal DNA is not correctly partitioned. In the same way, unbalanced DNA repartition might produce spores unable to germinate. Furthermore, the capacity to segregate equivalent DNA quantities in spores and consequently to produce viable spores is restored in strains that have lost the fusion state.

### Experimental procedures

#### Bacterial strains and culture conditions

*E. coli* SURE™ (Stratagene) strain was used as host for the construction of gene libraries and *E. coli* JM101 was used as a recipient for the M13mp18 derivatives and grown according to Sambrook *et al.* (1989).

*S. ambofaciens* DSM40697 wild-type (wt), mutant strains NSA27 and NSA65 (Leblond *et al.*, 1991) as well as derivatives isolated during this work were cultured at 30°C in YEME medium for PFGE experiments (Kieser *et al.*, 2000) or in Hickey-Tresner (HT; Pinnert-Sindico *et al.*, 1954) medium for standard DNA extraction. Surface cultures were carried out on HT solid medium for sporulation and germination assays and on minimum medium (MM) for DAPI-staining of mycelium and spores (Kieser *et al.*, 2000). All the spore suspensions used were filtered and titrated using a Toma cell (Kieser *et al.*, 2000). The germination rate consisted of the number of colony-forming units (cfus) after 4 days of growth divided by the microscopic spore counts.

Spontaneous derivatives of NSA27 and NSA65 were picked up at random in their progeny (marked NSA27<sub>S-number</sub> or NSA65<sub>S-number</sub>). Some derivatives, labelled NSA27<sub>r-number</sub>, were picked up in the progeny of a labelled NSA27 strain. The labelling consisted in site-specific integration (site *attB* of bacteriophage ΦC31) of the integrative plasmid pSET152 (Bierman *et al.*, 1992) or of a derivative of pSET152 harbouring the kanamycin resistance gene (instead of the original apramycin resistance gene cassette).

The levels of genomic instability reached in some mutant strains are so high that the genotype of derivatives has to be checked at each sporulation round. For that purpose, the derivatives whose PFGE *AseI* pattern was indistinguishable from the parental strain were considered as parental (P).

#### Plasmids

The gene libraries were constructed in cosmid Supercos1 (Evans *et al.*, 1989). Cloning of the 1.6 kb *PstI*-*SphI* fragment was cloned in the M13mp18 vector for sequencing.

© 2003 Blackwell Publishing Ltd, *Molecular Microbiology*

#### DNA manipulations

Extraction of total DNA from *S. ambofaciens* for the preparation of gene libraries, endonuclease analysis of high-molecular-weight genomic DNA in PFGE experiments (CHEF, Chu *et al.*, 1986; Bio-Rad), as well as purification of DNA fragments from gels, were carried out according to Leblond *et al.* (1996). Preparation of DNA samples for PFGE (with or without proteinase treatment) was achieved according to Leblond *et al.* (1996). *Saccharomyces cerevisiae* and *Hansenula wingei* chromosomes (Bio-Rad) were used as size markers in PFGE. DNA labelling and hybrid detection were carried out using the Dig DNA labelling system and detection kit (Roche Diagnostics). PCR reactions were carried out as described in Wenner *et al.* (2002) using consumables purchased from Roche Diagnostics in a Perkin Elmer thermal cycler. Nucleotide sequences were determined using the ABI Prism Dye Terminator Cycle Sequencing Kit and a 310 Prism automated DNA sequencer (Applied Biosystems). Oligonucleotides P<sub>618</sub> (5'CCCGCTCATCGACGAGAA3') and P<sub>267</sub> (5'CCCGCTGGAGCTGCTGAA3') were used to amplify the recombination point of NSA27 as well as to determine the wt sequences involved in the junction. PCR positive controls consisted in the amplification of a 330 bp portion of the *recA* sequence (Aigle *et al.*, 1997). Negative controls were carried out using water.

The *S. ambofaciens* NSA27 genomic library was constructed from a *Bam*HI partial digest of the total DNA in Supercos1, according to the recommendations of the supplier of the Gigapack III gold encapsidation kit (Stratagene). The wild-type library was constructed previously (Berger *et al.*, 1996). Cosmid DNA, double-stranded or single-stranded DNA of bacteriophage M13 derivatives were isolated as reported in Sambrook *et al.* (1989).

#### Electron microscopy

*Streptomyces* bacterial colonies were collected and fixed with 2% osmium tetroxide vapours for 72 h and then dehydrated by air-drying. Samples were then embedded in a matrix of EPON resin mixture and ultra-thin specimens were realized using an Ultracut E ultramicrotome. Specimens were further contrasted with uranyl acetate and lead nitrate. Electron micrographs were obtained using a Philips CM12 transmission electronic microscope.

#### DAPI-staining

Cultures were set up by inserting a sterile coverslip into the agar medium and spore inoculation along the glass surface. After a 2–3 days incubation at 30°C, cells were stained with DAPI (4',6-diamidino-2-phenylindole) as described (Flårdh *et al.*, 1999). Observations were performed using a microscope (Nikon) equipped for phase-contrast and epifluorescence microscopy. Photographs were taken with a digital camera (Nikon) and processed using Adobe Photoshop (Adobe 6.0) and Quantity One (Bio-Rad) softwares. Spore chains were chosen in the same focal plane in order to compare the relative intensity of spores within the same chain. Comparison of intensities of spores belonging to distinct spore chains was allowed by

relating the averages of all the spore chains to an arbitrarily chosen value (i.e. 600).

#### Statistical analyses

Probability values (*P*-values) are indicated when samples were significantly different from each other (Lamotte, 1971) at risks of 1% or 5%. The distributions of the intensities in DAPI-staining experiments were compared with the theoretical normal distribution using the mean and standard deviation of the sample tested (Kolmogorov–Smirnov test). When the distribution is normal, the observed and theoretical distributions were compared using the Fisher–Snédécór variance test.

#### Acknowledgements

T.W. was a fellow from the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie (MENRT), the recipient of a grant from the Association pour la Recherche sur le Cancer. V.R. was a fellow from the MENRT. A special thank to S. Gergalant for his help in computing the statistical tests. We thank Luc Marchal and Valérie Legué for their help in microscopy techniques and Ciaran Morrison for his help in the preparation of the manuscript.

#### References

- Aigle, B., Holl, A.C., Angulo, J.F., Leblond, P., and Decaris, B. (1997) Characterization of two *Streptomyces ambofaciens* *recA* mutants: identification of the RecA protein by immunoblotting. *FEMS Microbiol Lett* **149**: 181–187 (erratum **155**: 131–132).
- Altenbuchner, J., and Cullum, J. (1984) DNA amplification and an unstable arginine gene in *Streptomyces lividans* 66. *Mol Gen Genet* **195**: 134–138.
- Altenbuchner, J., and Cullum, J. (1985) Structure of an amplifiable DNA sequence in *Streptomyces lividans* 66. *Mol Gen Genet* **201**: 192–197.
- Aubert, M., Weber, E., Schneider, D., Simonet, J.M., and Decaris, B. (1993) Primary structure analysis of a duplicated region in the amplifiable AUD6 locus of *Streptomyces ambofaciens* DSM40697. *FEMS Microbiol Lett* **113**: 49–56.
- Bao, K., and Cohen, S.N. (2001) Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev* **15**: 1518–1527.
- Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* **417**: 141–147.
- Berger, F., Fischer, G., Kyriacou, A., Decaris, B., and Leblond, P. (1996) Mapping of the ribosomal operons on the linear chromosomal DNA of *Streptomyces ambofaciens* DSM40697. *FEMS Microbiol Lett* **143**: 167–173.
- Bidnenko, V., Ehrlich, S.D., and Michel, B. (2002) Replication fork collapse at replication terminator sequences. *EMBO J* **21**: 3898–3907.
- Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N., and Schoner, B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**: 43–49.
- Catakli, S., Andrieux, A., Leblond, P., Decaris, B., and Dary, A. (2003) Spontaneous chromosome circularization and amplification of a new amplifiable unit of DNA belonging to the terminal inverted repeats in *Streptomyces ambofaciens* ATCC23877. *Arch Microbiol* (in press). 1
- Chater, K.F., and Hopwood, D.A. (1984) *Streptomyces* genetics. In *The Biology of the Actinomycetes*. Goodfellow, M., Mordarski, M., and Williams, S.T. (eds). London: Academic Press, pp. 229–286.
- Chen, C.W., Yu, T.W., Lin, Y.-S., Kieser, H.M., and Hopwood, D.A. (1993) The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50 kb linear molecule. *Mol Microbiol* **7**: 925–932.
- Chu, G., Vollrath, D., and Davis, R.W. (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**: 1582–1585.
- Demuyter, P., Schneider, D., Leblond, P., Simonet, J.-M., and Decaris, B. (1991) A chromosomal region as a hotspot for multiple rearrangements associated with genetic instability in *Streptomyces ambofaciens* DSM40697. *J Gen Microbiol* **137**: 491–499. 2
- Dyson, P., and Schrempf, P. (1987) Genetic instability and DNA amplification in *Streptomyces lividans* 66. *J Bacteriol* **167**: 4796–4803.
- Evans, G.A., Lewis, K., and Rothenberg, B.E. (1989) High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* **79**: 9–20.
- Fischer, G., Decaris, B., and Leblond, P. (1997a) Occurrence of deletions associated with genetic instability in *Streptomyces ambofaciens*, is independent of the linearity of the chromosomal DNA. *J Bacteriol* **179**: 4553–4558.
- Fischer, G., Kyriacou, A., Decaris, B., and Leblond, P. (1997b) Genetic instability and its possible evolutionary implications on the chromosomal structure of *Streptomyces*. *Biochimie* **79**: 555–558.
- Fischer, G., Wenner, T., Decaris, B., and Leblond, P. (1998) Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proc Natl Acad Sci USA* **95**: 14296–14301.
- Flårdh, K., Findlay, K.C., and Chater, K.F. (1999) Association of early sporulation genes with suggested developmental decision points in *Streptomyces coelicolor* A3 (2). *Microbiology* **145**: 2229–2243.
- Hackett, J.A., Feldser, D.M., and Greider, C.W. (2001) Telomere dysfunction increases mutation rate and genomic instability. *Cell* **106**: 275–286.
- Huang, C.H., Lin, Y.S., Yang, Y.A., Huang, S.W., and Chen, C.W. (1998) The telomeres of *Streptomyces* chromosome contain conserved palindromic sequences with potential to form complex secondary structures. *Mol Microbiol* **28**: 905–916.
- Huang, J., Lih, C.J., Pan, K.H., and Cohen, S.N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev* **15**: 3183–3192.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M.,

- Kihuchi, H., Shiba, T., *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**: 526–531.
- Jakimowicz, D., Chater, K., and Zakrzewska-Czerwinska, J. (2002) The ParB protein of *Streptomyces coelicolor* A3 (2) recognizes a cluster of *parS* sequences within the origin-proximal region of the linear chromosome. *Mol Microbiol* **45**: 1365–1377.
- Kameoka, D., Lezhava, A., Zenitani, H., Hiratsu, K., Kawamoto, M., Goshi, K., *et al.* (1999) Analysis of fusion junctions of circularized chromosomes in *Streptomyces griseus*. *J Bacteriol* **181**: 5711–5717.
- Kieser, T., Bibb, M.J., Chater, K.F., and Hopwood, D.A. (2000) *Practical Streptomyces Genetics*. Norwich, UK: Foundation, J.I.
- Kim, H.J., Calcutt, M.J., Schmidt, F.J., and Chater, K.F. (2000) Partitioning of the linear chromosome during sporulation of *Streptomyces coelicolor* A3 (2) involves an *oriC*-linked *parAB* locus. *J Bacteriol* **182**: 1313–1320.
- Lemotte, M. (1971) *Initiation aux Méthodes Statistiques en Biologie*. Paris: Masson and Cieeds.
- Leblond, P., and Decaris, B. (1999) 'Unstable' linear chromosomes: the case of *Streptomyces*. In *Organization of the Prokaryotic Genome*. Charlebois, R.L. (ed.). Washington, DC: American Society for Microbiology, pp. 235–261.
- Leblond, P., Demuyter, P., Moutier, L., Laakel, M., Decaris, B., and Simonet, J.M. (1989) Hypervariability, a new phenomenon of genetic instability related to DNA amplification in *Streptomyces ambotaciens*. *J Bacteriol* **171**: 419–423.
- Leblond, P., Demuyter, P., Simonet, J.M., and Decaris, B. (1991) Genetic instability and associated genome plasticity in *Streptomyces ambotaciens*: pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. *J Bacteriol* **173**: 4229–4233.
- Leblond, P., Fischer, G., Franco, F.X., Berger, F., Guérineau, M., and Decaris, B. (1996) The unstable region of *Streptomyces ambotaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol Microbiol* **19**: 261–271.
- Lewis, P. (2001) Bacterial chromosome segregation. *Microbiology* **147**: 519–526.
- McClintock, B. (1951) Chromosome organization and genetic expression. *Cold Spring Harbor Symp Quant Biol* **16**: 13–47.
- Moore, J.K., and Haber, J.E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* **16**: 2164–2173.
- Pandza, K., Pfalzer, G., Cullum, J., and Hranueli, D. (1997) Physical mapping shows that the unstable oxytetracycline gene cluster of *Streptomyces rimosus* lies close the ends of the linear chromosome. *Microbiology* **143**: 1493–1501.
- Pang, X., Zhou, X., Sun, Y., and Deng, Z. (2002) Physical map of the linear chromosome of *Streptomyces hygroscopicus* 10–22 deduced by analysis of overlapping large chromosomal deletions. *J Bacteriol* **184**: 1958–1965.
- Pinner-Sindico, S., Ninet, L., Preud'Homme, J., and Cosar, C. (1954) A new antibiotic: spiramycin. *Antibiot Annu* **xxx** 4: 724–757.
- Qin, Z., and Cohen, S.N. (2000) Long palindromes formed in *Streptomyces* by nonrecombinational intra-strand annealing. *Genes Dev* **14**: 1789–1796.
- Qin, Z., and Cohen, S.N. (2002) Survival mechanisms for *Streptomyces* linear replicons after telomere damage. *Mol Microbiol* **45**: 785–794.
- Rauland, U., Glocker, I., Redenbach, M., and Cullum, J. (1995) DNA amplifications and deletions in *Streptomyces lividans* 66 and the loss of one end of the linear chromosome. *Mol Gen Genet* **246**: 37–44.
- Redenbach, M., Kleinert, E., and Stoll, A. (2000) Identification of DNA amplifications near the center of the *Streptomyces coelicolor* M145 chromosome. *FEMS Microbiol Lett* **191**: 123–129.
- Sakaguchi, K. (1990) Invertrons: a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of Adeno-type viruses. *Microbiol Rev* **54**: 66–74.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Uchida, T., Miyawaki, M., and Kinashi, H. (2003) Chromosomal arm replacement in *Streptomyces griseus*. *J Bacteriol* **185**: 1120–1124.
- Wenner, T., Roth, V., Decaris, B., and Leblond, P. (2002) Intragenomic and intraspecific polymorphism of the 16S-23S rDNA internally transcribed sequences of *Streptomyces ambotaciens*. *Microbiology* **148**: 633–642.
- Wheeler, R.T., and Shapiro, L. (1997) Bacterial chromosome segregation: is there a mitotic apparatus? *Cell* **88**: 577–579.

**PUBLICATION N°3**

Differential and crossed transcriptional expression of duplicated genes which encode alternative sigma factors in *Streptomyces ambofaciens*

Virginie Roth, Bertrand Aigle, Robert Bunet, Thomas Wenner, Céline Fourier, Bernard Decaris, and Pierre Leblond

*En préparation pour Journal of Bacteriology*

⋮

**Differential and crossed transcriptional control of duplicated genes encoding alternative sigma factors in *Streptomyces ambofaciens***

Virginie Roth, Bertrand Aigle, <sup>1</sup>Robert Bunet, <sup>2</sup>Thomas Wenner, Céline Fourier, Bernard Decaris, and Pierre Leblond\*

Laboratoire de Génétique et Microbiologie, UMR UHP-INRA 1128, IFR110, Faculté des Sciences et Techniques, Université Henri Poincaré, Nancy 1, Boulevard des Aiguillettes, BP 239, 54506 Vandoeuvre-lès-Nancy, France.

<sup>1</sup>Present address: Biological Institute, Department of Microbiology / Biotechnology, University of Tuebingen, Auf der Morgenstelle, Tuebingen, Germany.

<sup>2</sup>Present address: Chromosome Biology Lab, Department of Biochemistry, National University of Galway, University Road, Galway, Ireland.

\*Corresponding author :      Tél: 33 3 83 68 42 07  
   Fax: 33 3 83 68 44 99  
   Email: leblond@nancy.inra.fr

Key words: gene duplication, sigma factor, stress response, *Streptomyces*.

**6 Figures, 3 tables.**



## **Abstract**

The duplicated *hasR* and *hasL* genes encode alternative sigma factors (named  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$ ) belonging to the *B. subtilis* general stress response  $\sigma^{\text{B}}$  family factors. The duplication results from a recent event that occurred specifically in *Streptomyces ambofaciens*. The two genes share more than 98% of nucleotidic identity, and their deduced products exhibit more than 97% of amino acid identity. Contrasting with the coding sequences, their genetic environments as well as their transcriptional control strongly diverge. While *hasL* is monocistronic, *hasR* is preceded by two ORFs: *arsR* and *prsR* encoding putative anti-anti- $\sigma$  and anti- $\sigma$  factor respectively, and the three ORFs constitute a polycistronic unit.

Transcription of each gene is initiated from two promoters. The first one was shown to be developmentally controlled and to be similar to those recognised by the general stress response sigma factor  $\sigma^{\text{B}}$  of *B. subtilis*. Expression from these promoters dramatically increases in the course of the growth in liquid or solid media or following oxidative and saline stresses. It also appeared that *hasR* is 100 times more expressed than *hasL* from the  $\sigma^{\text{B}}$ -like promoter. In contrast, the both genes transcription from the second promoter (located upstream *arsR* in the case of the *hasR* locus) appears to be constitutive and weak.

Quantitative transcriptional analysis in single and double *has* mutant strains revealed that  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  direct their own transcription as well as that of the duplicate. No marked phenotype in response to growth conditions (media and stress) could be assigned to either single or double mutants revealing the probably redundancy of the  $\sigma$  factors and of the stress response systems.

## **Introduction**

Gene duplication is a key mechanism for genome evolution. This phenomenon was less frequently reported in bacteria than in eukaryotes. Gene duplication might allow the overproduction of a gene product. The most documented case is the duplication of the genes *tufA* and *tufB* which both appear to contribute to the synthesis of the elongation factor EF-Tu, a highly abundant cytoplasmic protein in *Escherichia coli* (15).

According to the Duplication-Degeneration-Complementation model (DDC) (14), the conservation of duplicated genes at the evolutive scale involves subfunctionalization of both members of the pair. Each duplicate can undergo complementary degenerative mutations involving loss or reduction or specialisation of its expression. The combined action of both loci would be necessary to assure the ancestral function. In *Streptomyces coelicolor* for example, while *glgBI* is involved in the glycogen metabolism during the vegetative growth phase, *glgBII* participates to the glycogen accumulation in spores and is only expressed in latest phases of differentiation (5).

According to Ikeda *et al.* (19), 35% of the *S. avermitilis* ORFs cluster into 721 paralogous families, with membership ranging from 2 to 91 genes per family. The membrane-spanning components of the ATP binding cassette (ABC) transporters, the two-component transcriptional regulator systems and the  $\sigma$  factors are the most spectacular examples of such gene families (19). The *S. coelicolor* (3) and *S. avermitilis* (19) genomes respectively contain 65 and 60  $\sigma$  factor encoding genes in contrast to 7 in *E. coli* (4), 16 in *B. subtilis* (23) and 13 in *M. tuberculosis* (8), the most related genus to *Streptomyces*. This characteristic might correlate with a complex life cycle associated to a high level of differentiation at both morphological (formation of aerial mycelium and sporulation) and biochemical (production of many secondary metabolites) levels, and the necessity to cope for environmental changes in their biotope (soil).

Nine genes, orthologues of the *B. subtilis* stress response  $\sigma^B$  factor, were found to form such a gene subfamily, in both *S. coelicolor* (3) (*sigF*, *sigG*, *sigH*, *sigI*, *sigJ*, *sigK*, *sigL*, *sigM*, *sigN*, according to G. Kelemen) and *S. avermitilis* (19), and to probably result from successive duplication events. The percentages of identity between the different deduced proteins of these nine genes ranged from 40% to 66%.

One members of this family, *sigJ* in *S. coelicolor* (also named *sigB* by Cho *et al.*, (7)) and SAV741 in *S. avermitilis*, was found in two copies in *S. ambofaciens* (13). The two copies were named *has* for homologous to alternative sigma factors, and present 98% of nucleotidic identity (13). Such a high level of identity between two duplicated genes has rarely been reported in a *Streptomyces* species (the *glgBI* and *glgBII* genes in *S. coelicolor* exhibit for example only 73% of nucleotidic identity). Clustering of the *S. coelicolor* ORFs by the Basic Local Alignment Search Tool protein clustering program (BLASTCLUST; minimum 95% identity, 100% length coverage) revealed indeed only ten couples of genes which might represent the same situation than the duplicated *has* genes of *S. ambofaciens*.

As *S. coelicolor* and *S. ambofaciens* are phylogenetically closely related, the duplication of the *has* genes seems to be recent and constitute an opportunity to study the fate of duplicated genes. For that purpose, the role of the *has* genes was investigated by transcriptional and mutational analyses.

## **Materials and Methods**

**Strains, plasmids and culture conditions.** *S. ambofaciens* DSM40697 was used through this study as wild-type strain. Strains and plasmids used are described in table 1. All *Streptomyces* culture conditions, media and selective conditions are described in (21). Luria-Bertani (LB) and SOB liquid media were used for growing *E. coli* and LB for *Bacillus subtilis* (31). Stress conditions were applied as follow : cells were grown to mid-exponential phase in liquid HT medium and exposed to the stressing agent for 30 minutes (10 mM H<sub>2</sub>O<sub>2</sub>, or 500 mM NaCl, or 34% saccharose or 0.7 M ethanol, or 30% cumene hydroperoxyde). High temperature shocks consisted in a shift from 30°C to 42°C for 20 minutes.

**Nucleic acid manipulations.** Standard or Pulsed-field gel electrophoresis (PFGE) DNA preparations from *S. ambofaciens* were carried out as described earlier (24, 25). Cosmid and plasmid DNAs were extracted from *E. coli* by the alkaline method (31). All restriction enzymes, PCR reagents and other molecular biology reagents were purchased from New England Biolabs or Roche Diagnostics.

DIG-DNA labelling (digoxigenin dUTP), hybridisation, washings and detection were done according to the supplier (Roche). Light emission was acquired with a Fluoro-S MultiImager (Bio-Rad). DNA sequencing was performed using the Dye terminator cycle sequencing kit on an ABI Prism 310 system and analysed with the Sequencing Analysis software (Applied Biosystem). All sequence similarity searches were performed in the nr database by using BLAST programs (2) at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/BLAST/>). Total RNA was isolated from HT liquid or solid (growth on cellophane) grown cultures of *S. ambofaciens* with the TRI Reagent<sup>TM</sup> (Sigma), followed by a chloroform step. High resolution S1 nuclease mapping experiments were performed according to Kieser *et al.* (21) using probes consisting of PCR products (sequences of primers and localisation, see table 2 and Figure 1).

**RT-PCR.** Five micrograms of total RNA were annealed to 2 µg of random hexamer primers (pdN6, Amersham Pharmacia Biotech Inc, USA) at 65°C for 5 minutes, then mixed with reverse transcription (RT) buffer (50 mM Tris-HCl, pH8.3, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM DDT), 20 µM desoxynucleoside triphosphates (Roche Diagnostics GmbH, Germany), 31.25 units of Guard (ribonuclease inhibitor, Amersham Pharmacia Biotech Inc, USA) and 10 units of DNaseI (Amersham Pharmacia Biotech Inc, USA) in a 25µl final volume. Reactions were carried out at 37°C for 20 minutes and stopped at 95°C for 5 minutes. Then, 200 units of MMLV Reverse Transcriptase (USB Corporation, USA) were added. Reactions were incubated at 37°C for 70 minutes and stopped at 95°C during 5 minutes. For the PCR step, 2 µl of cDNAs were amplified with different primer sets allowing the specific detection of *hasR* and *hasL* transcription. In parallel, two control PCRs (consisting of 2 µl of RNA taken out before and after the DNase step) were carried out to confirm the absence of DNA contamination in the RNA preparations.

**Real-Time PCR.** Real-time PCR was carried out on an iCycler iQ real-time PCR detection system (Bio-Rad) and the data were analysed using the software provided by the supplier. Sequences of primers and localisation are indicated in table 2 and Figure 1. Assays were performed using 5 µl of cDNA, 12.5 µl of Platinum Quantitative PCR SuperMix-UDG (Invitrogen), 3.3 µl SYBR Green I (10.000 × dilution, Sigma), and 10 pmol of each primer in a final volume of 25 µl. Thermal cycle conditions were as follow: 2 min at 50°C, 10 min at 95°C followed by 40 repeats maximum of 30 sec at 95°C and 1 min at 60°C (annealing and extension) and then 80 cycles of 10 sec with a temperature increasing by 0.5°C each cycle from 55°C to 94°C (melting curve). For each run, a standard dilution of the cDNA was used to check the relative efficiency of primers. A negative control (distilled water) was included in each real-time PCR run. Each sample was performed in duplicate. The *hrdB* gene was used as an internal control to quantify the relative expression of target gene. The expression levels of the downstream promoters (*hasRp1* and *hasLp1*) corresponded to the total gene expression

(quantified with Rqu1/Rqu2 and Lqu1/Lqu2, Figure 1 and table 2) deducted from those of the upstream promoters (*hasRp2* quantified with Rqu3/Rqu4 and *hasLp2* quantified with Lqu3/Lqu2, Figure 1 and table 2).

The Induction Factor (IF) corresponds to the difference of transcript amounts (quantified by real time RT-PCR) following the exposure to a stress relatively to a normal growth in exponential phase.

**PCR targeting mutagenesis.** The strategy used for gene disruption is the PCR-targeting system developed by Gust *et al.* (16). In order to replace a 839 bp sequence corresponding to the *hasR* coding region (included in cosmid 25E1), two primers, HasR1 and HasR2 (table 2 and Figure 1), which consisted of 39 nt sequence at the 5' end matching respectively the sequence adjacent to the start codon included and stop codon included of *hasR* and 20 nt /19 nt sequence at the 3' end matching each end of the disruption cassette were used to amplify the cassettes *aac3(IV)/oriT* and *aadA/oriT* from plasmids pIJ773 and pIJ778, respectively (16). All cassettes have the same ends and therefore could be amplified with the same set of primers. Similarly, the HasL1 and HasL2 primers (table 2 and Figure 1) were used to replace the 839 bp *hasL* sequence into the cosmid 14C4. Allelic exchanges were confirmed by Southern blot and PCR analysis.

**Complementation: construction of derivatives of pSET152s.** The coding sequence as well as promoters of each *has* gene were amplified using CoL1/CoL2 and CoR1/CoR2 sets of primers (table 2 and Figure 1) and cloned into the conjugative and integrative vector pSET152s (B. Gust, pers. com.) to give plasmids pSET152<sub>SL</sub> and pSET152<sub>SR</sub> respectively (table 1). The recombinant plasmids were sequenced in order to verify that no mutation in the *has* loci has occurred during the cloning process and were then introduced in the single mutant strains by conjugal transfer.

**Conjugation and screening of *S. ambifaciens* exconjugants.** Conjugation between *S. ambifaciens* and *E. coli* ET12567 (pUZ8002) containing the cosmid or plasmid of interest

were performed according to Kieser *et al.* (21). When needed, clones were replica-plated onto HT containing nalidixic acid, apramycin (or spectinomycin) with or without kanamycin (50  $\mu\text{g/ml}$ ) to select double crossover exconjugants ( $\text{Kan}^{\text{S}}$ , Apra<sup>-</sup>and/or Spec<sup>R</sup>).

## **Results and Discussion**

### **Characterisation of the *hasR* and *hasL* ORFs**

The nucleotide sequences of the duplicated *has* ORFs share 98% of identity over 846 pb. Each *has* gene is located on a different arm of the linear chromosome, *hasR* at about 450 kb from the right chromosomal end, and *hasL* at 850 kb from the left. The *has* genes constitute a substrate for homologous recombination and were discovered through their implication DNA rearrangements (13). The homologous area between the two *has* loci starts 4 nucleotides before the start codon and stops 8 nucleotides before the stop codon (Figure 1). The putative proteins,  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$ , only differ by eight amino acids over 281 residues, and share 100% identity in the regions 2-4 and 4-2 involved in the recognition of the -10 and -35 regions of promoters (26). The different homologues of  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  are presented in table 3. The  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  factors exhibit the highest similarity with  $\sigma^{\text{B}}$  of *S. coelicolor* (7); also named SigJ by Viollier *et al.*, (36). The genes controlled by  $\sigma^{\text{B}}$  were reported to play a role in osmoprotection and erection of aerial mycelium (7).

The *S. coelicolor* (3) and *S. avermitilis* (19) genomes sequences revealed the presence of a gene family of nine members, defined as homologues to  $\sigma^{\text{B}}$ , in which the orthologue of the *has* genes is *sigJ* (*sigB*) for *S. coelicolor* (94% / 93% of proteic identity with  $\sigma^{\text{BR}}$  /  $\sigma^{\text{BL}}$ ) and SAV741 for *S. avermitilis* (78% of proteic identity with  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$ ). The highest identity within this gene family concerns the products of the genes *sigJ* and *sigL* in *S. coelicolor*, and SAV741 and SAV1151 in *S. avermitilis*. These two protein couples exhibit 66% of identity, relatively to the 97% between  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  in *S. ambofaciens*. So, it appeared that the *has* gene orthologue is not duplicated in these two *Streptomyces* species.

Sequencing of the two *has* genes in the three strains of *S. ambofaciens* (DSM40697, ETH12567, ATCC23877) was also performed (data not shown). The identity between the two



*has* genes is quite similar for the 3 strains (98, 97 and 97% of nucleotidic identity between *hasR* and *hasL* in DSM40697, ETH12567, ATCC23877 strains respectively).

Since *S. avermitilis* is a species phylogenetically distant from *S. coelicolor* compared to *S. ambofaciens*, these data suggest that the duplication of the *has* genes is a recent event fixed in the evolution, which has occurred after the divergence between *S. coelicolor* and *S. ambofaciens* but before the divergence of the three strains of *S. ambofaciens*.

As shown in Figure 1, the sequenced region of *hasR* locus (4429 nt) contains five ORFs named *orfA*, *arsR*, *prsR*, *hasR* and *orfB*. Notably, *hasR* is preceded by two ORFs: *arsR* and *prsR* encoding respectively an anti-anti- $\sigma$  (91% of similarity with the putative anti-anti- $\sigma^B$  encoding by *rsbA* of *S. coelicolor*) and an anti- $\sigma$  factor (89% of similarity with the putative anti- $\sigma^B$  encoding by *rsbB* of *S. coelicolor*). The genetic organisation of the *hasR* locus is identical to that of *sigB* (*sigJ*) in *S. coelicolor* (Figure 2A). Interestingly, the conserved *sigB* (*sigJ*) locus is found on the left chromosomal arm of *S. coelicolor* at 640 kb from the DNA extremity while the *hasR* locus is located at 850 kb of the right extremity in *S. ambofaciens*. Thus, a large terminal rearrangement affecting at least 160 kb (overlapping the *hasR/sigB* (*sigJ*) locus) as revealed by PFGE experiments and hybridisation of recombinant cosmids (data not shown) had probably occurred since the divergence of the two species.

Further, the initiation codon of *prsR* overlaps the termination codon of *arsR* (as in *S. coelicolor*) suggesting a coregulation, as demonstrated below.

In contrast to the *hasR* locus, the genetic organisation of the *hasL* locus differs from that of *S. coelicolor* (Figure 2B). One ORF, *prsL*, encodes a protein showing only 56% of similarity with the product of *Sc5H1-15c*, a gene located on the right chromosomal arm of *S. coelicolor* at 580 kb from the chromosomal end and encoding a putative regulatory protein of anti- $\sigma$  factor-type. This gene lies downstream *sigL* (*Sc5H1-14*) in *S. coelicolor* (one of the nine  $\sigma$

factors of the family previously described and showing only 83% of similarity with  $\sigma^{BL}$ ), but in opposite orientation.

### Mapping of the Transcriptional Start Points (TSP)

For *hasR*, S1 nuclease mapping experiments using the P<sub>R1</sub> probe (Figure 1), consisting of 414 pb overlapping the intergenic region *hasR/prsR*, revealed two protected fragments (Figure 3A). The shortest one, F<sub>R1</sub> (102 nt), revealed a TSP located at 53 pb upstream the *hasR* start codon. The largest one (414 bp) could correspond either to probe reannealing or to a readthrough transcript from *arsR* promoter region (i.e. from *hasRp2*). RT-PCR experiments with the primers used for the construction of the P<sub>R1</sub> probe confirmed the existence of an operonic transcript (data not shown). These data are consistent with the organisation of the *hasR* locus suggesting that *hasR* was part of a three gene operon *arsR/prsR/hasR*, and an operonic organisation was also found for the *rsbA*, *rsbB* and *sigB* genes, homologues of the *hasR* locus in *S. coelicolor* (7). A second TSP (*hasRp2*) was indeed found 114 nt upstream from *arsR* using the probe P<sub>R2</sub> (fragment F<sub>R2</sub> of 157 nt, Figures 1 and 3A).

For *hasL*, two protected fragments of 92 nt (F<sub>L1</sub>) and 136 nt (F<sub>L2</sub>) were detected using the probe P<sub>L</sub> (446 pb, Figures 1 and 3A). They correspond to transcripts initiated from two TSP, *hasLp1* and *hasLp2*, localised respectively at 43 pb and 87 pb upstream the start codon.

Two boxes matching the consensus sequence of promoters recognised by the general stress response  $\sigma^B$  factor of *B. subtilis* (10) and  $\sigma^B$  of *S. coelicolor* (7) were identified upstream from *hasRp1* and *hasLp1* (Figure 4A). Upstream from the TSP *hasLp2*, a probable promoter for the principal  $\sigma$  factor,  $\sigma^{HrdB}$  of *S. coelicolor* (34) could be identified (Figure 4B). In contrast, only a degenerated  $\sigma^B$  promoter consensus was found upstream from *hasRp2* (Figure 4A). Interestingly, two sequences, almost identical to the promoters recognized by the sigma factor WhiG could be identified upstream *hasR* (Figure 4C). One of these putative promoters

(*hasRpWhiG1*) overlaps the *hasRp1* promoter. However, no transcription start site could be associated with these promoters in our experimental conditions, either with P<sub>R1</sub> probe or P<sub>R3</sub> probe (Figure 1). Furthermore, no transcript could be detected from surface differentiated mycelium although WhiG is involved in the late steps of cell differentiation (i.e. in the sporulation process, (6). It is important to note that the same sequence can be identified upstream *sigB* and overlapping the *sigBp1* promoter in *S. coelicolor* (Figure 4C).

### **The expression of the *has* genes is developmentally controlled**

S1 nuclease mapping experiments using RNAs prepared in the exponential (E), transition (T) and stationary (S) phases of growth showed that while the levels of *hasRp2* and *hasLp2* transcripts were rather constant, the transcription from *hasRp1* and *hasLp1* appeared to dramatically increase from the transition phase (Figure 3A). Real time RT-PCR analysis confirmed these observations (Figure 3B). The most remarkable increase was observed in the stationary phase where the *hasRp1* and *hasLp1* transcript abundance was 109 and 11 fold higher respectively than in exponential phase. When the levels of expression of the *has* genes were compared, it appeared that *hasR* is much more expressed than *hasL* with a maximum of 100 fold in stationary phase (Figure 3C). Analogous data were obtained using RNAs prepared from surface grown mycelium (data not shown).

The duplicated *has* genes seem to be subject to a complex transcriptional control with mainly two expression levels. Firstly, a basal level is ensured by two constitutive promoters (*hasLp2* and *hasRp2*) directing the transcription of *hasL* and of a polycistronic unit including *arsR-prsR-hasR* (anti-anti $\sigma$ /anti $\sigma$ / $\sigma$ ). Secondly, a developmentally controlled promoter was found upstream each *has* gene (*hasRp1* and *hasLp1*). Both promoter sequences would be recognised

by both  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  factors (auto- and cross-control) and by other  $\sigma$  factors probably of  $\sigma^{\text{B-}}$  type (see below).

### **The *has* genes belong to a stress response network**

Two approaches were used to determine the role of  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$ . Firstly, single and double *has* mutant strains were constructed in order to detect an eventual phenotype associated to the mutation. Secondly, transcriptional analyses in the wt and mutant strains were carried out to find growing conditions where the promoters detected previously might be induced.

The mutant strains were constructed using the PCR-targeting system described in Gust *et al.* (16). The single mutants were obtained by replacing a given gene (*hasR* or *hasL*) by a cassette including the apramycin resistance gene, and were named NSAR/NSAL for Nancy Streptomyces ambofaciens strains replaced for *hasR* / *hasL* (table 1). The double mutants were derived from the single mutants by replacing of the remaining *has* gene with a cassette harbouring the spectinomycin resistance gene. All the gene replacements were confirmed by southern hybridisation and PCR (see Material and Methods), and the absence of large chromosomal rearrangements was checked by PFGE (data not shown). Further experiments were carried out on three representative mutants of each type (table 1).

The single and double mutant strains were compared with the wt when grown on rich (HT, NA, SFM, R2) as well as on minimum media (MMG, MMM). No difference in growth rate or differentiation could be detected; nor could be detected difference when strains were grown at high temperature (37°C; instead of 30°C classically used) on solid media (HT, SFM, MMM or MMG), in acid or alkaline conditions (acid: pH 5.0 or alkaline: pH 9.0). The mutants showed no significant sensitivity to ultraviolet radiation exposure (0 to 400 J/m<sup>2</sup>). Furthermore, no growth impairment or phenotype could be noticed when exposed to osmotic stress conditions in liquid or solid media. Finally, a slight but reproducible sensitivity to

oxidative conditions distinguished the *has* mutant strains (single or double) from the wt strain. Using the disc paper method with H<sub>2</sub>O<sub>2</sub> concentrations ranging from 0.1 to 10 mM, the inhibition diameter of growth was 11.2% ( $\pm 5.6\%$ ) higher for the double mutant strains relatively to this of the wt strain (which varied of 2.8%). Moreover, preliminary experiments carried out on a  $\Delta prsR$  mutant strain (recently constructed by the PCR targeting system, (16), seemed to reveal a weak resistance of this strain to an oxidative stress (data not shown), confirming the potential rule of the  $\sigma^{BR}$  and  $\sigma^{BL}$  factors in this stress response.

The amounts of transcripts initiated at the different TSPs following stress treatments were then measured by real time PCR. Transcription from *hasRp1* and to a lesser extent from *hasLp1* appeared to be induced following the exposure to an oxidative stress (10 mM H<sub>2</sub>O<sub>2</sub>, induction factor (IF, see Material and Methods) 17 and 7.5, respectively) as well as to osmotic stress (500 mM NaCl, IF 25 and 21, respectively; 34% saccharose, IF 22.5 and 23, respectively). In the same experimental conditions, no induction was observed from *hasRp2* and *hasLp2* TSPs (IF ranging between 1 and 2.5). For each promoter, no increase of the transcript levels was observed after ethanol (4%), cumene hydroperoxyde (30%) or heat shock (42°C) treatments.

The same experiments were carried out in the mutant backgrounds. Surprisingly, it appeared that the IFs from *hasRp1* and *hasLp1* are not significantly different in the single and double mutants compared to the wt. Indeed, the IFs following an oxidative stress ranged between 13.2 and 16.5 for *hasRp1*, and between 6.9 and 9 for *hasLp1*, in the double (or single) mutant strains (17 and 7.5 respectively in the wt strain). The induction following an osmotic stress was also retained, as well after a salt treatment (IF ranged between 18 and 28.4 for *hasRp1*, and between 18 and 19.5 for *hasLp1* in the mutant strains, versus 25 and 21 in the wt strain); as after a treatment with 34% saccharose (IF of 24 for *hasRp1* and 18.2 for *hasLp1* in the double mutant strains, versus 22.5 and 23 in the wt strain).

These transcriptional analyses in the mutant strains showed that the transcriptional induction of the *hasRp1* and *hasLp1* promoters under stress conditions would result from transcription initiation by alternate  $\sigma$  factors.

These data reveal the probable redundancy of the  $\sigma$  factors involved in stress response in *Streptomyces*, as it has soon been demonstrated by the discovery of an absence of phenotype for the mutant strains interrupted for the *sigH* gene for example (36).

The sequence analysis of *S. coelicolor* genome (3) reveals a family of nine  $\sigma$  factors belonging to the  $\sigma^B$ -type. Among these, four have been isolated and reported so far ( $\sigma^F$ ,  $\sigma^G$ ,  $\sigma^H$  and  $\sigma^B$ ). The latter, also named SigJ (36), is the orthologue of the  $\sigma^{BR}$  and  $\sigma^{BL}$  factors, notably  $\sigma^{BR}$  as shown by the conservation of the genetic environment. Interestingly, our results differ from those obtained by Cho *et al.* (7) on  $\sigma^B$  of *S. coelicolor*. The *sigB* mutant strain is impaired in differentiation (deficient for aerial mycelium formation), and defence against osmotic stress (sensitive to hyperosmolarity). These phenotypical differences might be explained by a different transcriptional control of the *sigB* homologues, *hasR* and *hasL*.

Finally, the stress responses might rely on the activity of several specific  $\sigma$  factors and in particular for the osmotic stress response. Hence, in *S. coelicolor*, a sensory system that can co-ordinate the activity of multiple paralogous  $\sigma$  factors would be implied in the expression of genes belonging to the same stress regulon (36).

### **Transcriptional cross talk between the *has* genes.**

The quantification of the level of transcription from each TSP was achieved in the single and double mutant strains in the three growth phases in liquid medium.

All these data are resumed in Figure 5 and result from at least three independent real time RT-PCR experiments on each strain. Moreover, three independent single and double mutant

strains have been studied (table 1). Only the results concerning the *hasRp1* (Figure 5A) and *hasLp1* (Figure 5B) initiated transcripts will be detailed, since the rates of transcription from *hasRp2* (Figure 5C) and *hasLp2* (Figure 5D) are very low. All the transcription alterations exposed below are significant, considering the standard deviations of each value.

This analysis allowed to reveal a complex transcriptional control of the *has* genes with different types of regulation: (i) a positive auto-regulation; (ii) a cross-regulation; (iii) an inhibitory effect (probably by competition with other  $\sigma$  factors); (iv) a regulation by some  $\sigma^B$ -related sigma factor(s) of overlapping function. Moreover, it appeared that the effect of the  $\sigma^{BR}$  and  $\sigma^{BL}$  factors on each other was dependent on the growth phase (v).

#### (i) Positive auto-regulation

The transcript level from *hasRp1* decreased 30 fold in a  $\Delta hasR$  mutant in stationary phase of growth and 1.9 fold in transition phase (Figure 5A). These data indicated that the expression of *hasR* gene from *hasRp1* is under the control of its own product ( $\sigma^{BR}$ ). This effect was not observed in the exponential phase of growth (see iii). In the same way, *hasL* was also self-regulated as revealed in stationary phase of growth, since *hasLp1* transcript level decreased 1.8 fold in a  $\Delta hasL$  mutant.

#### (ii) Cross-regulation

In stationary phase of growth, the transcription level from *hasLp1* decreased by 19 fold in the  $\Delta hasR$  mutant strain (Figure 5B). This suggested that the transcription of *hasL* might be directed by  $\sigma^{BR}$  in the wt. The cross-regulation was also observed in transition phase (2 fold decrease) but remained not detected in exponential phase of growth (see iii).

Reciprocally, *hasR* may be subject to a cross regulation by  $\sigma^{BL}$ . In a  $\Delta hasL$  mutant, a slight but significant decrease of the *hasRp1* transcription was indeed noticed in exponential, transition and stationary phases of growth with a 1.4, 1.2, and 2.6 fold decreases, respectively (Figure 5A).

To confirm that the altered transcriptional levels truly resulted from the  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  deficiency, complementation experiments were realised and strains were constructed in which a wt copy of *hasR* or *hasL* under the control of its own promoter was introduced into the NSAR or NSAL strains respectively, using the integrative vector pSET152s (see Material and Methods). Transcription rates were nearly fully restored in the complemented strains. As an example, the transcription levels from *hasLp1* and *hasRp1* were restored at 74% and 87% respectively, in a complemented  $\Delta\textit{hasR}$  strain (Figures 5A and 5B).

### (iii) Inhibitory effect

In the  $\Delta\textit{hasL}$  mutant strains, the transcription rate from *hasLp1* increased by 26 fold (exponential phase, Figure 5B) compared to the wt strain, suggesting  $\sigma^{\text{BL}}$  might have a negative effect on the transcription of its own gene. The same inhibitory effect was observed in transition phase with a 3.6 fold increase in the  $\Delta\textit{hasL}$  mutant (Figure 5B). Similarly, the  $\sigma^{\text{BR}}$  factor seems to exert an inhibitory effect on its own promoter in exponential phase of growth (2.9 fold, Figure 5A). This inhibitory effect might also be crossed in exponential phase of growth, since the deficiency in  $\sigma^{\text{BR}}$  is accompanied by a slight increase of the *hasLp1* transcript level (2.4 fold, Figure 5B).

This inhibitory effect can result from the competition of two (or more)  $\sigma$  factors for the same promoter or for the RNA polymerase core enzyme. The latter effect was characterized for  $\sigma^{\text{S}}$  and  $\sigma^{70}$  in *E. coli* competing for a limiting amount of RNA polymerase core enzyme during stationary phase (12). It has also been suggested in the case of the SigH, SigI and SigJ  $\sigma$  factors implied in the osmotic stress response in *S. coelicolor* (36).

The ability to bind to the core RNAP depends of their relative affinities for core RNAP, as well as the relative abundance of each  $\sigma$  factor. In our model, the abundance of each free  $\sigma^{\text{BR}}$



and  $\sigma^{BL}$  would depend of the anti-anti- $\sigma$ /anti- $\sigma$  control system found in the vicinity of the *has* genes.

Another level of competition could result from the presence of promoters recognised by the sigma factor WhiG upstream *hasR*, one of them overlapping the *hasRp1* promoter (Figure 4). The binding of WhiG (efficient or not) might prevent the recognition of *hasRp1* by  $\sigma^{BR}$  by a simple steric effect.

(iv) A  $\sigma$  network including the  $\sigma^{BR}$  and  $\sigma^{BL}$  factors

It is important to note the additional effects of the  $\sigma^{BR}$  and  $\sigma^{BL}$  deficiency in the double mutants  $\Delta hasR\Delta hasL$ . When a transcriptional induction was observed in each single mutant strain, this induction was indeed higher in the double mutant strain (case of *hasLp1* in exponential phase, Figure 5B). Similarly, when opposite effects were observed in each single mutant strain (cases of *hasRp1* in exponential phase, Figure 5A; and of *hasLp1* in transition phase, Figure 5B), the alteration of the transcription in the double mutant was mitigated. Finally, when the deficiency of each  $\sigma^{BR}$  or  $\sigma^{BL}$  factor was accompanied by a reduction of the transcriptional level, this decrease was higher in the double mutant (cases of *hasRp1* in transition or stationary phases, Figure 5A; and of *hasLp1* in stationary phase, Figure 5B).

Moreover, the detection of these residual transcriptions in the double mutants  $\Delta hasR\Delta hasL$  allowed to suggest that other  $\sigma$  factor(s) might regulate the transcription of the *has* genes. This evidence of  $\sigma$  factor redundancy is in accordance with the maintenance of the *has* genes transcriptional induction under stress conditions in the mutant strains (see above). These data are also coherent with the results of Viollier *et al.* (36), suggesting the existence of shared promoter recognition specificity within the stress response  $\sigma$  factors.

(v) Growth phase dependence

We can represent all these results as three global models, one for each growth phase, of *has* genes regulation at the transcriptional level, but without considering the *hasRp2* and *hasLp2* initiated transcripts which are responsible for the basal expression of *has* genes (Figures 6).

Of course, the complexity of this regulation model can reach a higher degree if we consider the post transcriptional regulation level with the anti- $\sigma$  factors which can inhibit the  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  activity.

As shown previously, the cross-control exerted on both *hasRp1* and *hasLp1* promoters exhibit different patterns depending of the growth phase. Thus, the positive control exerted at a given phase can be negative at another stage.

For example, when  $\sigma^{\text{BR}}$  is responsible for its positive regulation in transition and stationary phases, it has a negative effect in exponential phase. The control of *hasL* by  $\sigma^{\text{BR}}$  shows the same type of switch with an inhibition in exponential phase and a positive effect in both transition and stationary phases.

Similar controls were observed for  $\sigma^{\text{BL}}$  on its own promoter *hasLp1* with a negative effect in exponential and transition phases and an activation in stationary phase. In contrast, a positive cross regulation of  $\sigma^{\text{BL}}$  on *hasRp1* was observed in all three phases.

The absence of negative effect of the  $\sigma^{\text{BR}}$  and  $\sigma^{\text{BL}}$  in the latest phase of growth might be due to their inhibition by the anti- $\sigma$  factor.

The different regulation levels of the *has* genes in *S. ambofaciens* are in accordance with previous studies in *S. coelicolor*, particularly those concerning genes encoding  $\sigma$  factors. Firstly, many genes are under the control of multiple promoters. For example, four promoters have been characterised in the case of the *sigH* gene, belonging to the  $\sigma^{\text{B}}$  family (20, 22). As for the *has* promoters, these four promoters are differentially expressed at various stress conditions and stages of differentiation. Moreover, the autoregulation, usual for the bacterial

stress response  $\sigma$  factors (17), has also been demonstrated in the case of *sigH* (20, 32), as for the *hasR* genes.

Finally, post-transcriptional regulation by a anti-anti- $\sigma$ /anti- $\sigma$  control system was demonstrated or suspected for SigH (33, 37) as for several other  $\sigma$  factors including, for example,  $\sigma^R$ , implied in the oxidative stress response (28).

## Acknowledgements.

V.R. was a fellow from the Ministère de l'Education nationale, de la Recherche et de la Technologie (MENRT). T.W. was a fellow from the MENRT, the recipient of a grant from the Association pour la Recherche sur le Cancer.

We thank JIC and in particular B. Gust, K. Chater, and T. Kieser for their help in the construction of the mutant strains.

## References

1. **Ainsa, J. A., H. D. Parry, and K. F. Chater.** 1999. A response regulator-like protein that functions at an intermediate stage of sporulation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **34**:607-19.
2. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.
3. **Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141-7.
4. **Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-74.
5. **Bruton, C. J., K. A. Plaskitt, and K. F. Chater.** 1995. Tissue-specific glycogen branching isoenzymes in a multicellular prokaryote, *Streptomyces coelicolor* A3(2). *Mol Microbiol* **18**:89-99.
6. **Chater, K. F., C. J. Bruton, K. A. Plaskitt, M. J. Buttner, C. Méndez, and J. D. Helmann.** 1989. The developmental fate of *S.coelicolor* hyphae depends upon a gene product homologous with the motility  $\sigma$  factor of *B.subtilis*. *Cell* **59**:133-143.

7. **Cho, Y. H., E. J. Lee, B. E. Ahn, and J. H. Roe.** 2001. SigB, an RNA polymerase sigma factor required for osmoprotection and proper differentiation of *Streptomyces coelicolor*. *Mol Microbiol* **42**:205-14.
8. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. Mclean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537-44.
9. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**:6640-5.
10. **Engelmann, S., C. Lindner, and M. Hecker.** 1995. Cloning, nucleotide sequence, and regulation of *katE* encoding a sigma B-dependent catalase in *Bacillus subtilis*. *J Bacteriol* **177**:5598-605.
11. **Evans, G. A., K. Lewis, and B. E. Rothenberg.** 1989. High efficiency vectors for cosmid microcloning and genomic analysis. *Gene* **79**:9-20.
12. **Farewell, A., K. Kvint, and T. Nystrom.** 1998. Negative regulation by RpoS: a case of sigma factor competition. *Mol Microbiol* **29**:1039-51.
13. **Fischer, G., T. Wenner, B. Decaris, and P. Leblond.** 1998. Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *Proceedings of the National Academy of Sciences USA*:14296-14301.
14. **Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait.** 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**:1531-45.
15. **Furano, A. V.** 1978. Direct demonstration of duplicate *tuf* genes in enteric bacteria. *Proc Natl Acad Sci U S A* **75**:3104-8.
16. **Gust, B., G. L. Challis, K. Fowler, T. Kieser, and K. F. Chater.** 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **100**:1541-6.
17. **Hecker, M., W. Schumann, and U. Volker.** 1996. Heat-shock and general stress response in *Bacillus subtilis*. *Mol Microbiol* **19**:417-28.
18. **Hutter, R.** 1967. [Classification of the streptomycetes with special regard to the antibiotics formed from them]. *Bibl Microbiol* **6**:1-382.

19. **Ikeda, H., J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, and S. Omura.** 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**:526-31.
20. **Kelemen, G. H., P. H. Viollier, J. Tenor, L. Marri, M. J. Buttner, and C. J. Thompson.** 2001. A connection between stress and development in the multicellular prokaryote *Streptomyces coelicolor* A3(2). *Mol Microbiol* **40**:804-14.
21. **Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood.** 2000. *Practical Streptomyces genetics*. The John Innes Foundation, Norwich.
22. **Kormanec, J., B. Sevcikova, N. Halgasova, R. Knirschova, and B. Rezuchova.** 2000. Identification and transcriptional characterization of the gene encoding the stress-response sigma factor sigma(H) in *streptomyces coelicolor* A3(2). *FEMS Microbiol Lett* **189**:31-8.
23. **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Düsterhöft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, A. Fabret, E. Ferrari, B. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, J. Golightly, E. Grandi, F. Guiseppi, J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, B. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R. P. Mellado, M. Mizuno, A. Moestl, S. Nakai, M. Noback, B. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, M. Prescott, E. Presecan, P. Pujic, A. Purnelle, et al.** 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**:249-56.
24. **Leblond, P., P. Demuyter, J. M. Simonet, and B. Decaris.** 1991. Genetic instability and associated genome plasticity in *Streptomyces ambofaciens* : pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. *J Bacteriol* **173**:4229-4233.
25. **Leblond, P., G. Fischer, F. X. Francou, F. Berger, M. Guérineau, and B. Decaris.** 1996. The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. *Mol Microbiol* **19**:261-271.
26. **Lonetto, M., M. Gribskov, and C. A. Gross.** 1992. The  $\sigma$ 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**:3843-3849.
27. **MacNeil, D. J.** 1988. Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. *J Bacteriol* **170**:5607-12.

28. **Paget, M. S., V. Molle, G. Cohen, Y. Aharonowitz, and M. J. Buttner.** 2001. Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol Microbiol* **42**:1007-20.
29. **Ray, C., R. E. Hay, H. L. Carter, and C. P. Moran, Jr.** 1985. Mutations that affect utilization of a promoter in stationary-phase *Bacillus subtilis*. *J Bacteriol* **163**:610-4.
30. **Ryding, N. J., G. H. Kelemen, C. A. Whatling, K. Flårdh, M. J. Buttner, and K. F. Chater.** 1998. A developmentally regulated gene encoding a repressor-like protein is essential for sporulation in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **29**:343-357.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New-York.
32. **Sevcikova, B., O. Benada, O. Kofronova, and J. Kormanec.** 2001. Stress-response sigma factor sigma(H) is essential for morphological differentiation of *Streptomyces coelicolor* A3(2). *Arch Microbiol* **177**:98-106.
33. **Sevcikova, B., and J. Kormanec.** 2002. Activity of the *Streptomyces coelicolor* stress-response sigma factor sigmaH is regulated by an anti-sigma factor. *FEMS Microbiol Lett* **209**:229-35.
34. **Strohl, W. R.** 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Research* **20**:961-974.
35. **Tan, H., and K. F. Chater.** 1993. Two developmentally controlled promoters of *Streptomyces coelicolor* A(3)2 that resemble the major class of motility-related promoters in other bacteria. *J Bacteriol* **175**:933-940.
36. **Viollier, P. H., G. H. Kelemen, G. E. Dale, K. T. Nguyen, M. J. Buttner, and C. J. Thompson.** 2003. Specialized osmotic stress response systems involve multiple SigB-like sigma factors in *Streptomyces coelicolor*. *Mol Microbiol* **47**:699-714.
37. **Viollier, P. H., A. Weihofen, M. Folcher, and C. J. Thompson.** 2003. Post-transcriptional regulation of the *Streptomyces coelicolor* stress responsive sigma factor, SigH, involves translational control, proteolytic processing, and an anti-sigma factor homolog. *J Mol Biol* **325**:637-49.

## Tables

TABLE 1. Bacterial strains and plasmids used in this study.

Strain	Property	Reference
<b><i>S. ambofaciens</i></b>		
DSM40697	Wild-type	(18)
NSAR1, NSAR2, NSAR3	<i>hasR</i> replaced by <i>aac(3)IV/oriT</i> cassette	This study
NSAL1, NSAL2, NSAL3	<i>hasL</i> replaced by <i>aac(3)IV/oriT</i> cassette	This study
NSARL1, NSARL3	<i>hasR</i> and <i>hasL</i> replaced by <i>aac(3)IV/oriT</i> and <i>aadA/oriT</i> cassettes, respectively	This study
NSALR2, NSALR7	<i>hasL</i> and <i>hasR</i> replaced by <i>aac(3)IV/oriT</i> and <i>aadA/oriT</i> cassettes, respectively	This study
NSAR2-s1, NSAR2-s2	NSAR2 with pSET152s integrated into the chromosomal <i>attB</i> site of $\phi$ C31	This study
NSAR2-s <sub>R</sub> 1, NSAR2-s <sub>R</sub> 2	NSAR2 with pSET152 <sub>sR</sub> integrated into the chromosomal <i>attB</i> site of $\phi$ C31	This study
<b><i>E. coli</i></b>		
DH5 $\alpha$	General cloning strain	(31)
ET12567	Used for intergeneric conjugation	(27)
BW25113	Used for PCR targeting mutagenesis	(9)
<b>Plasmid/Cosmid</b>		
14C4	Cosmid (containing <i>hasL</i> locus)	(13)
25E1	Cosmid (containing <i>hasR</i> locus)	(13)
pIJ790	<i>gam</i> , <i>bet</i> , <i>exo</i>	(16)
pIJ773	<i>oriT</i> , <i>aac(3)IV</i>	(16)
pIJ778	<i>oriT</i> , <i>aadA</i>	(16)
pSET152s	<i>oriT</i> , <i>attP</i> , <i>int</i> , <i>aadA</i>	B. Gust, pers.com.
pSET152 <sub>sR</sub>	pSET152s + <i>hasR</i> cloned into the <i>EcoRI</i> site	This study
Supercos1	<i>bla</i> , <i>neo</i>	(11)

*gam*, Gam (host's exonucleaseV inhibitor) gene; *bet*, Bet (ssDNA binding protein) gene; *exo*, exonuclease gene; *aadA*, spectinomycin and streptomycin resistance gene; *aac(3)IV*, apramycin resistance gene; *neo*, kanamycin resistance gene; *bla*, ampicillin resistance gene; *int*, integrase gene of  $\phi$ C31; *oriT*, origin of transfer.



**TABLE 2. Primers used in this work.**

Construction of the mutant strains: the sequence in bold corresponds to the *hasR* or *hasL* locus sequence (including the start and stop codons which are underlined) while the sequence in normal characters matches the end of the disruption cassette.

	Primer	Nucleotide sequence (5'-3')	Use
Construction of mutant strains	HasR1	GCCGTGAGCGGGCCGATCGACTGGGAGGAACATCATG <u>TGTAGCCTGGACCTCCTTC</u>	Replacement of <i>hasR</i>
	HasR2	ACCGCCGCGGGAAGGACAGCCGCCCTGAGCGGCGCTCAATCCGGGGATCCGTCGACC	Replacement of <i>hasR</i>
	HasL1	GTGACGAGCAGGCAGTTCGAACAGGGAGGGCCCATCATG <u>TGTAGGCTGCAGCTGCTTC</u>	Replacement of <i>hasL</i>
	HasL2	GCCGCCCGGTGTGCCACGGTGAGGGGCCCTCGCCTCAATCCGGGGATCCGTCGACC	Replacement of <i>hasL</i>
Complementation	CoR1	TATGAATTCTGTGACGGTGTGCCGATG	Complementation of <i>hasR</i> mutant
	CoR2	AATGAATCCCGGGCGGTGAGGATGA	Complementation of <i>hasR</i> mutant
	CoL1	TATGAATTCGACCGGTGCGTGTGAGGC	Complementation of <i>hasL</i> mutant
	CoL2	ATTGAATTCGCCGACGTCTCCCGCGCC	Complementation of <i>hasL</i> mutant
Real time RT-PCR	R <sub>qu</sub> 1	CCGATGAGGCCGGCCGT	Gene expression of <i>hasRp1</i> and <i>hasRp2</i>
	R <sub>qu</sub> 2	GATGTTCCCTCCAGTCGATCGG	Gene expression of <i>hasRp1</i> and <i>hasRp2</i>
	R <sub>qu</sub> 3	CCGACGAGACTGCGCTTAC	Gene expression of <i>hasRp2</i>
	R <sub>qu</sub> 4	GGCGAGACAGACCGAGAAAA	Gene expression of <i>hasRp2</i>
	L <sub>qu</sub> 1	CTGCTGACGAGCAGGCAG	Gene expression of <i>hasLp1</i> and <i>hasLp2</i>
	L <sub>qu</sub> 2	GATGGGCCCTCCCTGTT	Gene expression of <i>hasLp1</i> and <i>hasLp2</i>
	L <sub>qu</sub> 3	CCCGAGGCCGGGGTATTG	Gene expression of <i>hasLp2</i>
	HrdB-R	CGCGAGCCCATCTCGCTGC	Gene expression of <i>hrdB</i>
HrdB-F	CGTCGAGGGTCTTCGGCTG	Gene expression of <i>hrdB</i>	
Probes for SI mapping	R5	GTCATCCACGGTGGTGGCAGCGCA	Probes for <i>hasR</i> (P <sub>R1</sub> , P <sub>R3</sub> )
	G	GTCCCTGGGCGCCACCTTGGACGGG	Probe for <i>hasR</i> (P <sub>R3</sub> )
	E	GGCATGCCCGTGGTGGTGTGCTCGG	Probes for <i>hasR</i> and <i>hasL</i> (P <sub>L</sub> , P <sub>R1</sub> )
	L5	CACGCACAGCGTGTACGTGGTGTCC	Probe for <i>hasL</i> (P <sub>L</sub> )
	arsR1	CGGTGTCAACGTAGACGGA	Probe for <i>hasR</i> (P <sub>R2</sub> )
	arsR2	TCTCCACTGGCCCGGATAC	Probe for <i>hasR</i> (P <sub>R2</sub> )

**TABLE 3. Homologues of  $\sigma^{BR}$  and  $\sigma^{BL}$ .**

	<i>S. a.</i> $\sigma^{BR}$	<i>S. a.</i> $\sigma^{BL}$
<i>S. a.</i> $\sigma^{BL}$	97% (100% - 100%)	
<i>S. c.</i> $\sigma^B$ (SigJ)	93% (100% - 100%)	92% (100% - 100%)
<i>S. c.</i> SigL	86%(95.5% - 74%)	83% (95.5% - 74%)
<i>S. c.</i> CrtS	64% (100% - 62.5%)	63% (100% - 62.5%)
<i>S. c.</i> SigF	47% (73% - 62.5%)	46% (73% - 62.5%)
<i>B. s.</i> $\sigma^B$	34% (73% - 50%)	33% (73% - 50%)

For each comparison, the first percentage represents the similarity between the two deduced proteins. The percentages of identity of the regions 2-4 and 4-2, respectively, are given into brackets.

## **Figure legends**

### **Figure 1: Genetic organisation of *hasR* and *hasL* loci.**

The hatched rectangle represents the homologous area between the two *has* loci. The ribosome binding sites (RBS) are symbolised by the small black squares. The transcription start sites (*hasRp1*, *hasRp2*, *hasLp1* and *hasLp2*) are marked by bent arrows. *orfA* is incompletely sequenced (dotted arrow).

S1 probes ( $P_{R1}$ ,  $P_{R2}$ ,  $P_{R3}$  and  $P_L$ ) are presented as black bars with asterisks for the radiolabelled ends. The dotted thick features symbolised the protected fragments obtained with each S1 probe.

The different primers used in real time RT-PCR experiments, construction of S1 probes or for the genetic constructions are symbolised by the open arrows (see table 2).

### **Figure 2: Comparison of *has* loci organisation between *S. ambofaciens* and *S. coelicolor*.**

The percentages correspond to the similarity between the deduced products of the genes. The intensity of the different gray boxes is proportional to the level of similarity.

The 86% of similarity between the products of *orfA* (incompletely sequenced, dotted arrow) and *ScF55-21c* concerned the 104 first amino acids of the proteins.

The double bar symbolises the non conservation of the *hasL* locus organisation since the *Sc5E9-03* ORF is located 80 kb upstream from *sigL* in *S. coelicolor*.

\*: according to Cho *et al.*, (7).

\*\* : according to Viollier *et al.*, (36).

### **Figure 3: Identification and quantification of the four *has* transcripts.**

**A.** Identification of transcription start sites of *hasR* (on left) and *hasL* (on right) genes by S1 nuclease mapping using P<sub>R2</sub>, P<sub>R1</sub> and P<sub>L</sub>, probes respectively, and determination of the level of transcription of the *has* genes from liquid growth culture. Total RNA were prepared from the wt strain at the exponential (E, after 9 hours of growth), transition (T, after 16 hours of growth) and stationary (S, after 24 hours of growth) phases of growth in liquid HT medium. Only the part of the autoradiographic film corresponding to each protected fragment (F<sub>R1</sub>, F<sub>R2</sub>, F<sub>L1</sub> and F<sub>L2</sub>) is shown. These experiments have been made with three independent RNA samples for each growth phase.

**B.** Quantification of the two *hasR* transcripts (on left) and of the two *hasL* transcripts (on right) in the growth course in liquid HT medium by real time RT-PCR (see Materials and Methods, *hrdB* was used as an internal reference as it is expressed at a constant level during the growth). The level of expression of each transcript in the exponential phase was arbitrarily fixed at 1. These experiments have been made with the same three independent RNA samples for each growth phase than those used for S1 mapping experiments.

**C.** Comparison of the levels of expression of each *has* transcript in the course of the growth in liquid HT medium by real time RT-PCR. The level of expression in the exponential phase from the less expressed transcript - corresponding to *hasLp2* initiated transcript - was arbitrarily fixed at 1.

#### Figure 4: Promoter sequences.

A. Alignment of the *hasRp1*, *hasLp1* and *hasRp2* promoters with  $\sigma^B$ -type promoters : Pctc of *B. subtilis* (29) and *sigBp1* promoter of *S. coelicolor* (7).

B. Alignment of the *hasLp2* promoter with the promoter consensus recognized by the principal sigma factor  $\sigma^{\text{HrdB}}$  in *S. coelicolor* (34).

C. Alignment of the two putative WhiG-type promoters identified upstream *hasR* coding region and of this identified upstream *sigB* (*sigJ*) coding region with promoters recognised by WhiG in *S. coelicolor*. P*whiH* and P*whiI* are the promoter sequences of the *whiH* and *whiI* genes known to be regulated by WhiG in *S. coelicolor* (30; 1). PTH4 and PTH270 are two sequences identified as targets of  $\sigma^{\text{WhiG}}$  in *S. coelicolor* (35).

The nucleotides in bold are those in common with the promoter consensus sequence. The two nucleotides underlined (TA) in the -10 region of *hasRp1* (or *sigBp1*) promoter and in the -35 region of *hasRpWhiG1* (or *sigBpWhiG*) promoter are the same and symbolised the overlapping of these promoters.

**Figure 5: Quantification in the growth course of transcription levels from *hasRp1* (A.), *hasLp1* (B.), *hasRp2* (C.), and *hasLp2* (D.) in the wt, single and double mutant strains and complemented strains.**

The mutant genetic contexts are named  $\Delta hasR$ ,  $\Delta hasL$  or  $\Delta hasR\Delta hasL$  and result from the analysis of three independent strains of each type (see table 1). The complementation experiments correspond to the  $\Delta hasR+pSETsR$  data and result from the analysis of two independent strains (see table 1).

For each strain, three independent RNA samples have been analysed by real time RT-PCR (allowing calculating an average and a standard deviation).

The values are relative to the level of expression from *hasLp2* in exponential phase, arbitrarily fixed at 1.

**Figure 6: Models of *has* genes regulation in exponential (A.), transition (B.) and stationary (C.) phases of growth deduced from the real time RT-PCR experiments.**

The thickness of the arrows is proportional to the level of regulation (quantified by real-time RT-PCR, see Figure 5).

**Figure 1.**

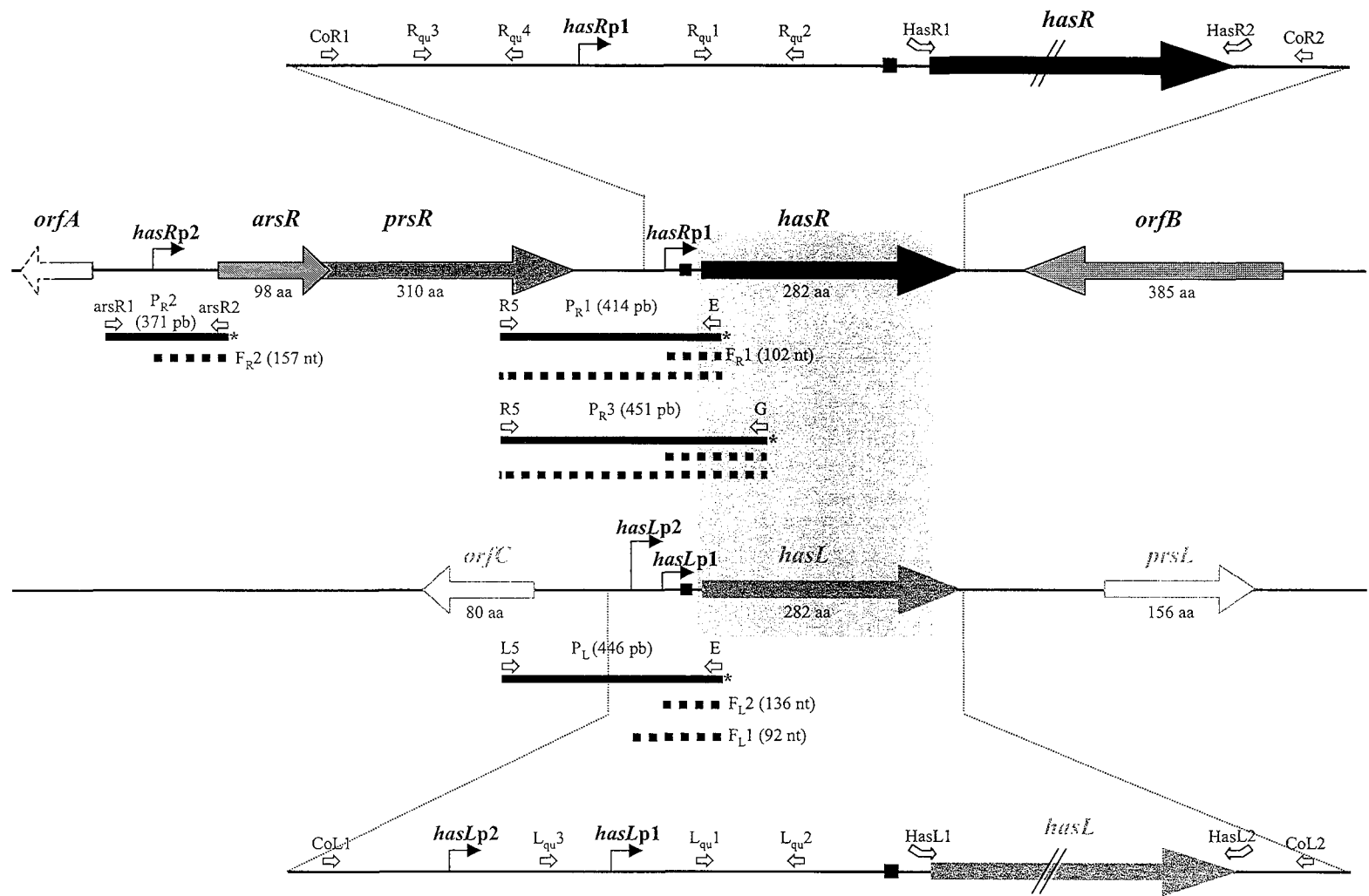
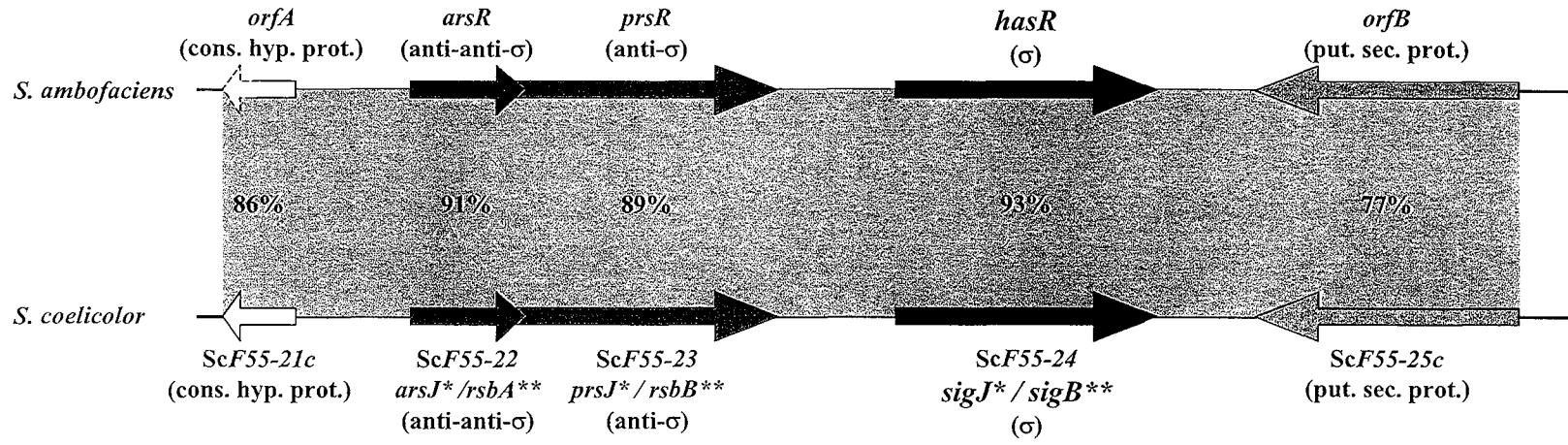


Figure 2.

A.



B.

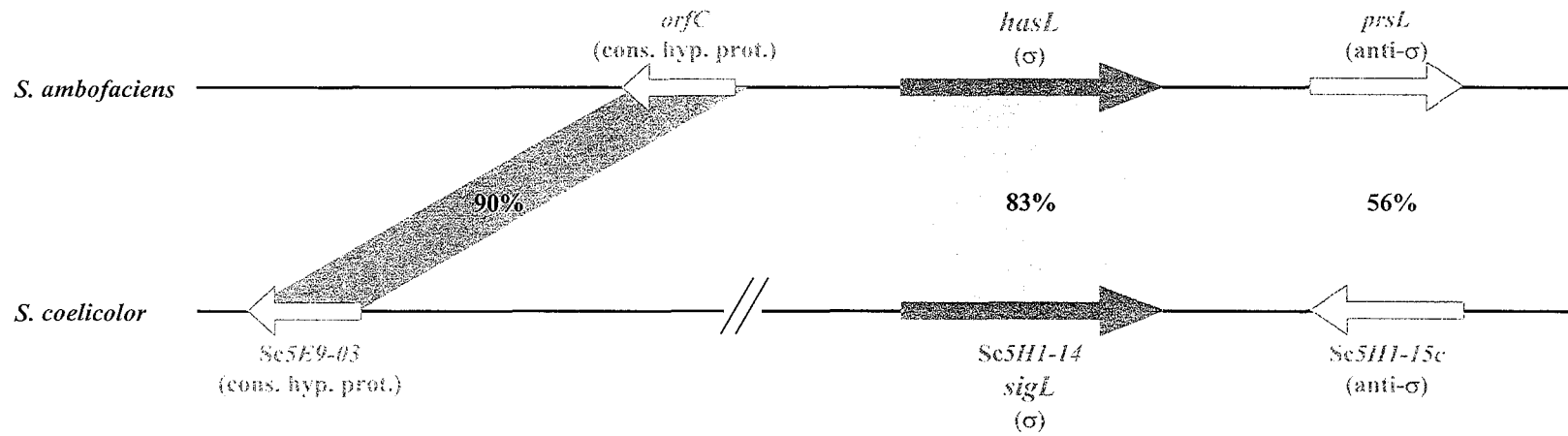




Figure 3.

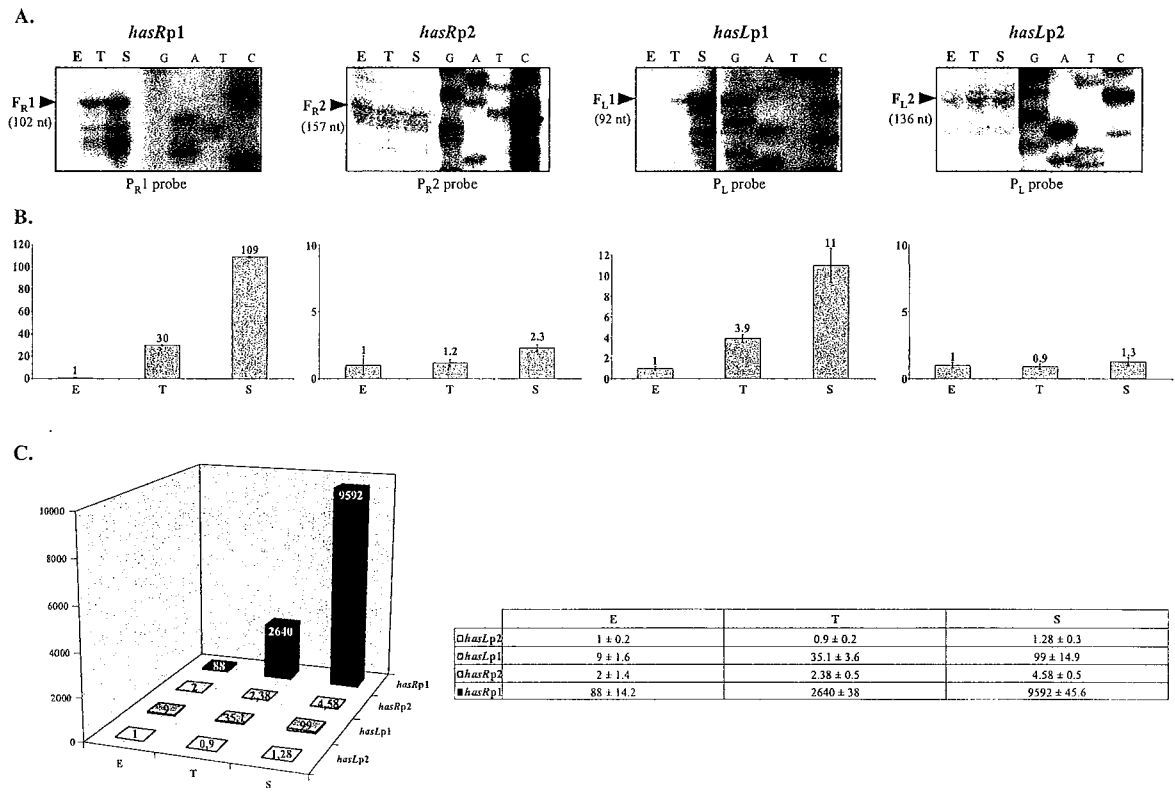


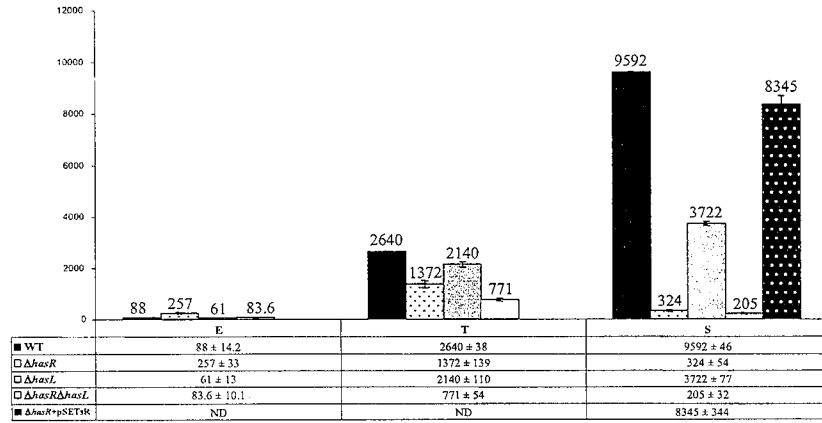
Figure 4.

		-35		-10	+1
<b>A.</b>	consensus	<b>anGnnTnn</b>	-N <sub>12/14</sub> -	<b>GGGtA</b>	
	<i>Pctc</i>	<b>AGGTTTAA</b>	-N <sub>14</sub> -	<b>GGGTA</b>	-N <sub>11</sub> - A
	<i>sigBp1</i>	<b>AGGTGTGC</b>	-N <sub>13</sub> -	<b>GGG<u>T</u>A</b>	-N <sub>7</sub> - G
	<i>hasRp1</i>	<b>ATGCATGA</b>	-N <sub>13</sub> -	<b>GGG<u>T</u>A</b>	-N <sub>11</sub> - G
	<i>hasLp1</i>	<b>TTGCATAC</b>	-N <sub>13</sub> -	<b>GGGTA</b>	-N <sub>9</sub> - T
	<i>hasRp2</i>	<b>GGGCGTAC</b>	-N <sub>12</sub> -	<b>CGGTA</b>	-N <sub>10</sub> - C

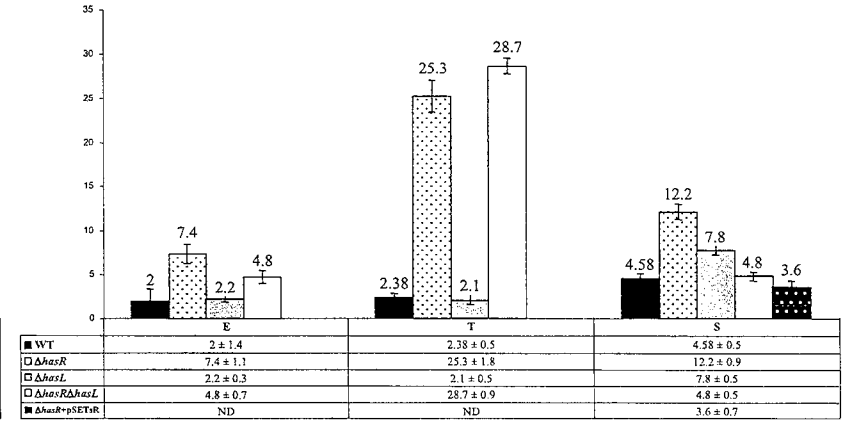
		-35		-10	+1
<b>B.</b>	consensus	<b>TTGACR</b>	-N <sub>16/18</sub> -	<b>TAgRRT</b>	
	<i>hasLp2</i>	<b>TGGAGG</b>	-N <sub>17</sub> -	<b>TATGAA</b>	-N <sub>7</sub> - C

		-35		-10
<b>C.</b>	consensus	<b>TAAA</b>	-N <sub>14/16</sub> -	<b>GCCGATAA</b> <sub>/T</sub>
	<i>PwhiH</i>	<b>TCAA</b>	-N <sub>16</sub> -	<b>GCCGACAA</b>
	<i>PwhiI</i>	<b>TAAA</b>	-N <sub>16</sub> -	<b>GCCGAAGA</b>
	<i>PTH4</i>	<b>TGGA</b>	-N <sub>16</sub> -	<b>GCCGAACT</b>
	<i>PTH270</i>	<b>TACG</b>	-N <sub>15</sub> -	<b>CCCGATGA</b>
	<i>hasRpWhiG1</i>	<b><u>TAAG</u></b>	-N <sub>13</sub> -	<b>GCCGATGA</b>
	<i>sigBpWhiG</i>	<b><u>TAGG</u></b>	-N <sub>13</sub> -	<b>GCCGACAA</b>
	<i>hasRpWhiG2</i>	<b>TGAG</b>	-N <sub>15</sub> -	<b>GCCGATCG</b>

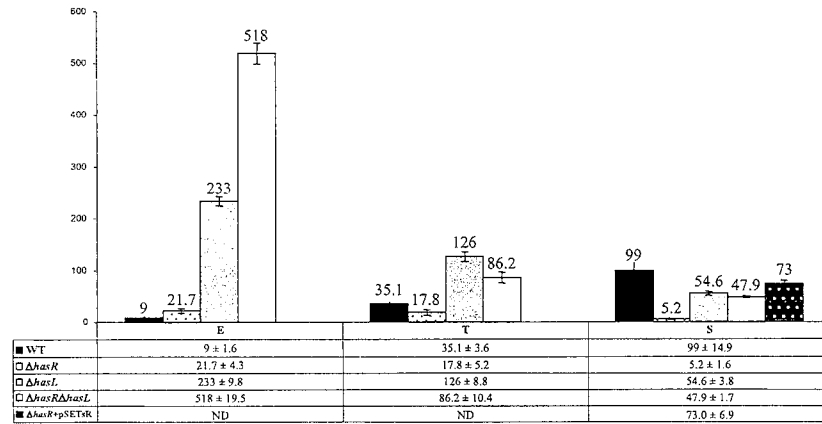
**A. *hasRp1***



**C. *hasRp2***



**B. *hasLp1***



**D. *hasLp2***

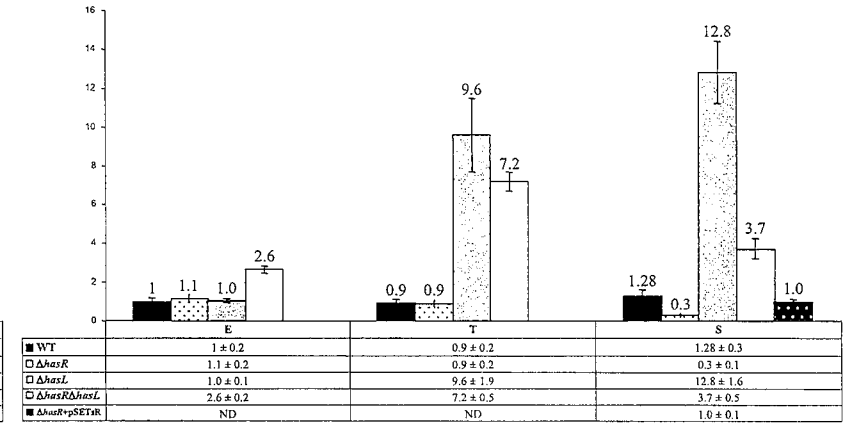
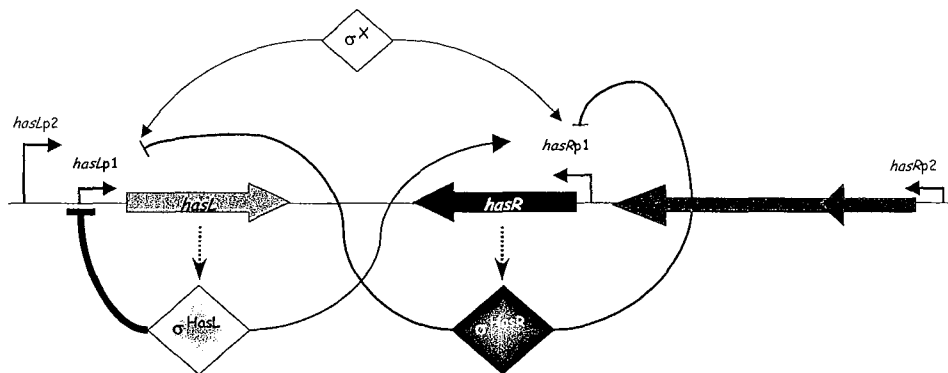
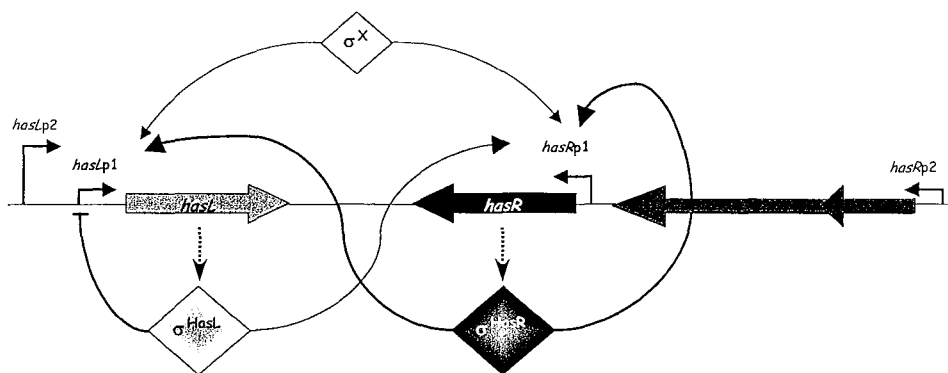


Figure 6.

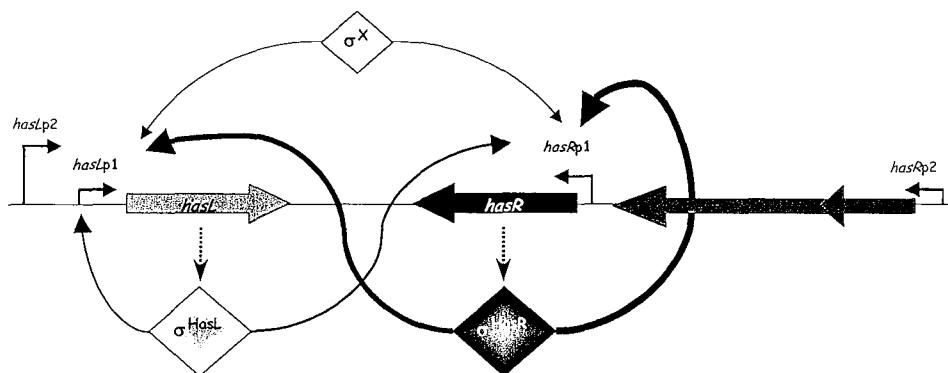
A. Exponential phase of growth



B. Transition phase of growth



C. Stationary phase of growth





## RESUME

Les réarrangements chromosomiques constituent un moteur pour l'évolution de la structure chromosomique et de l'information génétique. Chez *Streptomyces ambofaciens* l'instabilité génétique est étroitement corrélée à la formation de grands réarrangements génomiques, affectant les régions subtélomériques du chromosome linéaire.

La bactérie *Streptomyces ambofaciens* peut être considérée comme un modèle d'étude des phénomènes d'évolution des génomes par duplication du matériel génétique, et ce aussi bien en termes dynamiques qu'en termes évolutifs.

☞ La dynamique du phénomène de duplication peut en effet être abordée grâce à la caractérisation des réarrangements présents chez certains mutants issus de l'instabilité génétique de *S. ambofaciens*. L'analyse des descendances de mutants, porteurs d'une fusion chromosomique, a permis de montrer que cette structure conférait un état « mutateur », à l'origine d'une grande diversité chromosomique. Des duplications géniques de grande ampleur sont notamment générées aux extrémités du chromosome.

☞ Une duplication génique, localisée dans les régions terminales du chromosome, a été impliquée dans la formation d'un grand réarrangement génomique (Fischer *et al.*, 1998). Ces gènes dupliqués *hasR* et *hasL* (pour « Homologous to Alternative Sigma factors », et R et L pour leur localisation respective sur le bras chromosomique droit et gauche) coderaient des facteurs  $\sigma$  alternatifs, homologues du facteur  $\sigma^B$  de réponse générale aux stress chez *Bacillus subtilis*.

L'analyse du polymorphisme des régions intergéniques des six opérons codant les ARN ribosomiques (ITS 16S-23S) chez *S. ambofaciens*, *S. coelicolor* et *S. griseus*, a permis de confirmer les relations phylogénétiques entre ces différentes espèces. *S. coelicolor* A3(2), phylogénétiquement proche de *S. ambofaciens*, possède une unique ORF fortement homologue des ORF *has*, et ne présente donc pas la duplication génique décelée chez *S. ambofaciens*. La duplication des gènes *has* chez *S. ambofaciens* est donc un événement récent en terme évolutif, et pourrait résulter d'un réarrangement chromosomique, à l'image des réarrangements de l'instabilité génétique.

L'impact de cette duplication sur l'information de l'espèce *S. ambofaciens* a été recherché par deux approches. La construction de mutants, visant à associer un phénotype à la mutation de l'une voire des deux copies des gènes *has*, a permis de montrer que ces facteurs sigma participaient à la réponse générale aux stress (notamment oxydatif et osmotique). Une analyse transcriptionnelle fine (par RT-PCR quantitative), a mis en évidence l'appartenance de ces deux facteurs  $\sigma$  à un réseau de régulation complexe, avec notamment des phénomènes d'autorégulation, de régulation croisée, d'inhibition mais aussi avec l'intervention d'autres facteurs  $\sigma$ , soulignant la redondance de ces facteurs transcriptionnels chez les *Streptomyces*.