

UNIVERSITE DE LILLE1
SCIENCES ET TECHNOLOGIES

THESE

Présentée par **ARTHUR TAPI**

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE DE LILLE1, SCIENCES ET TECHNOLOGIES
Ecole Doctorale Sciences de la Matière, du Rayonnement et de l'Environnement
Option : **Ingénierie des Fonctions Biologiques**

**Stratégie moléculaire de mise en évidence de peptides actifs
d'origine non ribosomiale chez *Bacillus sp.* et *Lactobacillus sp.***

Préparée au Laboratoire ProBioGEM, UPRES EA 1026
Polytech'Lille, France

Soutenue le 21 Janvier 2010 devant le Jury composé de :

Philippe JACQUES , Professeur (Laboratoire ProBioGEM, Polytech'Lille1, France)	Directeur
Patrick FICKERS , Professeur (Université Libre de Bruxelles, Belgique)	Rapporteur
Sigrig FLAHAUT , Professeur (Université Libre de Bruxelles, Belgique)	Rapporteur
Marlène CHOLLET , Maître de Conférences (Laboratoire ProBioGEM, Lille 1, France)	Examinatrice
Catherine FONTAGNE-FAUCHER , Maître de Conférences (LBAE-IUT, Toulouse, France)	Examinatrice
Maude PUPIN , Maître de Conférences (Laboratoire d'Informatique Fondamentale, Lille1, France)	Examinatrice

Avant-propos

Cette thèse a été pour moi plus qu'un parcours académique, elle aura été un alibi pour une aventure personnelle.

Au nom de cette thèse, j'ai rencontré des personnes de divers horizons que je n'imaginais pas lors de mon arrivée cette année-là, en plein hiver à Zaventem après avoir parcouru 7000 Km.

Ce manuscrit est le résultat de plusieurs années d'études allant des sciences des aliments à la biologie moléculaire en passant par la bioinformatique. Cette thèse a nécessité mon passage dans plusieurs universités (Université d'Abobo-Adjamé, Faculté des Sciences Agronomiques de Gembloux et Université des Sciences et Technologies de Lille).

Ces travaux sont le fruit entre la collaboration de plusieurs chercheurs issus de laboratoires différents (Laboratoire Wiame De Bruxelles et de Chimie Biologie de Gembloux en Belgique, les laboratoires de Procédés Biologiques, Génie Enzymatique et Microbien et d'Informatique Fondamentale de Lille en France).

Ce manuscrit est un condensé de plusieurs approches à l'interface de plusieurs disciplines (biomoléculaires, bioinformatiques, de biotechnologies et de biochimie).

Cette thèse a permis la publication d'un article scientifique dans une revue internationale, de deux autres articles soumis et en cours de publication, d'un congrès international, de communications orale et sous forme d'affiches, puis enfin d'un logiciel libre.



Remerciements

Je tiens tout d'abord à exprimer ma reconnaissance et toute ma gratitude au Professeur Philippe Jacques, enseignant-chercheur à Polytech'lille, France, qui a dirigé cette thèse.

Merci à vous pour la confiance, le soutien permanent et la liberté nécessaire dont j'ai bénéficié de votre part durant ces travaux. Vos conseils avisés, votre disponibilité permanente à toutes mes sollicitations ont été un atout indéniable pour l'aboutissement de cette thèse. J'éprouve une énorme fierté d'avoir appris auprès de vous. J'ose espérer que ceci n'est que le début d'une collaboration future.

Ma profonde gratitude à Mon pays qui m'a soutenu financièrement tout au long de mes études supérieures. Je remercie tout particulièrement le Colonel Mirassou d'avoir exaucé mon souhait de terminer mes études supérieures. J'adresse mes sincères sentiments de reconnaissance au Cdt Amaury et à Mr. Amédée Couassi-Blé.

Mes sentiments distingués à l'endroit du Docteur Marlène Chollet, qui n'a ménagé aucun moment, ni son quotidien dans la mise en œuvre de ce travail. Cela a été un enthousiasme constant d'apprendre et de progresser à tes côtés durant ces années. Merci Marlène pour ces années passées ensemble à s'inquiéter d'un résultat, à essayer de le comprendre et à l'interpréter. Merci pour la confiance et la complicité que tu as su installer, cela a été sans nul doute un apport indéniable pour l'aboutissement de ces travaux. J'espère qu'on aura encore des projets à venir.

Je tiens tout particulièrement à remercier Mr. Bart Scherens qui m'a initié dans la bioinformatique et continué à me guider dans mes difficultés tout au long de ces années. Merci Bart d'avoir cru si tôt en ces travaux et en moi.

J'y associe tout le personnel de Wiame pour la convivialité durant mon séjour d'études.

Ma profonde gratitude au Docteur Maude Pupin d'avoir accepté que je m'insère dans ces cours pour mieux appréhender mon sujet de thèse. Merci Maude pour vos conseils. Je ne suis pas surpris que vous ayez accepté d'évaluer mes travaux de thèse.

Je remercie les Professeurs Sigrid Flahaut et Patrick Fickers de m'avoir fait l'honneur d'être les rapporteurs de cette thèse. J'éprouve un profond respect pour vos travaux et parcours. Le regard critique et avisé que vous avez porté sur mes travaux ne peut que m'encourager à être encore plus perspicace et engagé dans mes recherches.

J'adresse mes vives remerciements aux Professeurs Didier Guillochon et Pascal Dhulster qui ceux sont succédés à la Direction du Laboratoire Procédés Biologiques, Génie Enzymatique et Microbien durant ces travaux.

A tous les enseignants-chercheurs et personnels de ProbioGEM. En particulier, à Max Béchet qui n'a compté ni son temps ni ses conseils pour mes manipulations parfois nocturnes. Merci Max pour cette disponibilité et cette bonne humeur permanente. A toi Cathy pour toute ta sollicitude et ta disponibilité.

Aux Professeurs Georges Amani et Koissy-Kpein Michel qui m'ont encouragé à entamer cette aventure et permis sa réalisation académique. A vous, je ne cesserai d'être reconnaissant.

Aux Personnels de l'UFR-STA de l'Université d'Abobo-Adjamé et à ces différents responsables.

A toi ma fiancée Lucie, merci pour ton soutien indéfectible et permanent. Dieu seul sait les moments pénibles qui ont jalonné ces années de recherche. Tu as été là, à chaque instant, à mes moments de grands doutes mais aussi de joies. J'espère que très bientôt tu présenteras aussi ta thèse.

A toute ma famille, Sylvain Tapi, Hiltrud Zylka, Marcellin et Carolle Loba, Sylvain Gogo, Boni Gnadja, Davide Naguegnon, aux familles Gnaoré et Goba, à mes frères et sœurs, à tous mes proches. Merci de vos encouragements et votre soutien indéfectible malgré ces longues années très loin de vous.

Les mots ne suffiront pas pour vous dire combien de fois j'ai espéré vous voir et partager des petits moments comme par le passé.

A tous ceux qui ont participé à ces travaux, au Dr. Aly Savadogo, à Ahmed Abderahmani, à tous les stagiaires en microbiologie et en bioinformatique avec qui j'ai travaillé. Merci d'avoir supporté mes humeurs, mes angoisses, parfois mes doutes et par moments mes coups de gueules. J'ai dans l'ensemble avec chacun passé des moments de bonheur et de complicité. Bien merci à tous. Cette thèse n'est qu'une petite partie de tout ce qu'on a fait ensemble.

Ce travail n'aurait pas été possible sans les nombreux sacrifices que vous avez bien voulu consentir.

A tous mes compagnons de thèse de Belgique et de France (Philomène, Koné, Eric, Marius, Privat, Frédéric, Venant, Aurelien, Mathieu, François, Estelle, Omar, Parvin, Joany, Francine, Souleymane, Franck, Amani, Maxime, Jean Sebastien, Laurent, Sameh, Walaa, Layal, Angélique). J'en oublie certainement...

A tous ceux qui de près ou de loin ont contribué à l'aboutissement de ces travaux.

Dédicace

A ma fille âgée de moins 3 mois,

A ma grand-mère,

Table des Matières

Liste des figures.....	5
Liste des Tableaux.....	8
Liste des abréviations.....	10
INTRODUCTION GENERALE	12
CHAPITRE 1 LES PEPTIDES NON RIBOSOMIAUX	15
I. HISTORIQUE DES PEPTIDES NON RIBOSOMIAUX	15
II. DIVERSITES ET ACTIVITES BIOLOGIQUES DES PEPTIDES NON RIBOSOMIAUX	16
III. LE MECANISME DE LA SYNTHÈSE NON RIBOSOMIALE	20
3.1. <i>Structure modulaire des NRPS</i>	20
3.1.1. Domaines principaux	20
3.1.2. Domaines secondaires	24
3.2. <i>Biosynthèse de molécules non ribosomiales</i>	28
3.2.1. La biosynthèse linéaire : cas de l'opéron bacitracine chez <i>Bacillus</i>	28
3.2.2. La biosynthèse non-linéaire	29
3.2.3. Description de la biosynthèse itérative	30
IV. LES EXCEPTIONS A LA SYNTHÈSE NON RIBOSOMIALE CLASSIQUE	31
V. STRUCTURE MODULAIRE DES POLYKETIDES SYNTHASES.....	32
5.1. <i>Polykétides Synthases de type I</i>	32
5.2. <i>Polykétides Synthases de type II</i>	34
5.3. <i>Polykétides Synthases de type III</i>	34
5.4. <i>Les hybrides NRPS/PKS</i>	35
VI. LES DIFFÉRENTES STRATÉGIES D'IDENTIFICATION DES MOLECULES NON RIBOSOMIALES	35
6.1. <i>Approches moléculaires</i>	35
6.2. <i>Approches prédictives bioinformatiques</i>	38
6.2.1. Outils permettant la détermination du substrat activé	39
6.2.2. Outils permettant la détermination de l'agencement structural des opérons	39
CHAPITRE 2 LES LIPOPEPTIDES: DES PEPTIDES NON RIBOSOMIAUX DE BACILLUS	41
I. STRUCTURES ET ACTIVITES BIOLOGIQUES DES LIPOPEPTIDES	41
II. ORGANISATION DES OPERONS QUI CODENT LES SYNTHETASES DE LIPOPEPTIDES	44
2.1. <i>Opéron des surfactines</i>	45
2.2. <i>Opéron des fengycines</i>	45
2.3. <i>Opéron des iturines</i>	46
CHAPITRE 3 BUT DU TRAVAIL	48
CHAPITRE 4 MATERIELS ET METHODES.....	50
I. MATERIELS.....	50
1.1. <i>Milieux de culture</i>	50
1.1.1. Milieu de Man–Rogosa–Sharpe ou MRS.....	50
1.1.2. Milieu de Luria-Bertani ou LB	50
1.1.3. Milieu DSM	50
1.1.4. Milieu Potato Dextrose Agar ou PDA.....	50
1.1.5. Milieux 853 et 856	51
1.1.6. Milieux 863 et 868	51
1.1.7. Milieu SOC	51
1.1.8. Milieu nutritif.....	51
1.1.9. Eau peptonée.....	51

1.2.	<i>Souches bactériennes</i>	51
1.2.1.	Souches de <i>Lactobacillus</i>	51
1.2.2.	Souches de <i>Bacillus</i>	52
1.2.3.	Autres souches	54
1.3.	<i>Vecteurs plasmidiques</i>	54
1.4.	<i>Outils bioinformatiques</i>	56
1.4.1.	Logiciels d'analyses	56
1.4.2.	Sites Webs	57
II.	METHODES	59
2.1.	<i>Techniques moléculaires</i>	59
2.1.1.	Extraction de l'ADN	59
2.1.2.	Préparation des échantillons pour PCR sur colonies	59
2.1.3.	Conditions expérimentales de la PCR	60
2.1.4.	Electrophorèse	61
2.1.5.	Purification des acides nucléiques	61
2.1.6.	Ligation et clonage	62
2.1.7.	Transformation et sélection des colonies	62
2.1.8.	Extraction de plasmides	63
2.1.9.	Séquençage	63
2.2.	<i>Techniques bioinformatiques</i>	63
2.2.1.	Détermination des amorces à partir des gènes NRPS de <i>L. plantarum</i>	64
2.2.2.	Détermination des amorces pour la détection de gène NRPS de <i>Bacillus</i>	65
2.3.	<i>Identification des souches</i>	68
2.3.1.	Identification moléculaire	68
2.3.2.	Identification biochimique	68
2.3.3.	Détermination d'activités biologiques	69
2.3.4.	Caractérisation des souches isolées des condiments fermentés africains	71
2.4.	<i>Recherche de lipopeptides non ribosomiaux</i>	72
2.4.1.	Mesure de la tension de surface	72
2.4.2.	Recherche de lipopeptides par HPLC	72
2.4.3.	Recherche de lipopeptides par Maldi-TOF/LC-MS/MS-MS	73
CHAPITRE 5 DEFINITION D'AMORCES SPECIFIQUES POUR L'AMPLIFICATION DE GENES NON RIBOSOMIAUX PAR PCR A PARTIR DES GENES NRPS DE <i>LACTOBACILLUS</i>		75
I.	ELABORATION DES AMORCES A PARTIR DES GENES NRPS DE <i>LACTOBACILLUS</i>	75
1.1.	<i>Organisation des gènes NRPS de Lactobacillus</i>	75
1.1.1.	La définition des amorces	76
1.1.2.	Calcul du Coefficient d'Hybridation	79
1.2.	<i>Application sur des souches modèles du genre Lactobacillus</i>	80
1.2.1.	Tests PCR <i>in vitro</i>	80
1.2.2.	Séquençage de fragments amplifiés par PCR	83
II.	CRIBLAGE DE BACTERIES LACTIQUES POUR LA RECHERCHE DE MOLECULES ACTIVES NON RIBOSOMIALES	86
2.1.	<i>Recherche d'activités biologiques</i>	86
2.2.	<i>Recherches de gènes NRPS dans la collection de bactéries lactiques</i>	89
	<i>Conclusion et perspectives</i>	92
CHAPITRE 6 DETECTION PAR PCR DES GENES NON RIBOSOMIAUX CODANT POUR LA BIOSYNTHESE DE LIPOPEPTIDES DE <i>BACILLUS SP.</i>		93
I.	STRATEGIE DE LA MISE AU POINT DES AMORCES SPECIFIQUES	93
1.1.	<i>Comparaison des opérons</i>	93
1.2.	<i>Définition des amorces</i>	95
1.3.	<i>Calcul du Coefficient d'hybridation (HyC)</i>	96
II.	VALIDATION DES AMORCES	97

2.1.	<i>Tests PCR in silico</i>	97
2.2.	<i>Tests PCR in vitro</i>	98
	<i>Conclusions et Perspectives</i>	103
CHAPITRE 7 IDENTIFICATION DE SOUCHES DU GENRE <i>BACILLUS</i> PRODUCTRICES DE LIPOPEPTIDES D'ORIGINE NON RIBOSOMIALE ISOLEES DE DEUX PRODUITS FERMENTES AFRICAINS : LE SOUMBALA ET LE BIKALGA.....		106
I.	IDENTIFICATION DES SOUCHES ISSUES DES PRODUITS FERMENTES.....	106
1.1.	<i>Isolement des souches</i>	106
1.2.	<i>Identification des souches isolées</i>	109
1.2.1.	Tests biologiques et physico-chimiques	109
1.2.2.	Caractéristiques morphologiques et biochimiques des souches	111
1.2.3.	Tests moléculaires d'identification	112
II.	RECHERCHE DE LIPOPEPTIDES NON RIBOSOMIAUX.....	115
2.1.	<i>Analyses biochimiques</i>	115
2.1.1.	Tension de surface des surnageants de culture.....	115
2.1.2.	Recherche de lipopeptides par Maldi-TOF	116
2.2.	<i>Recherche de gènes de lipopeptides synthétases</i>	119
	<i>Conclusions et Perspectives</i>	122
CHAPITRE 8 MISE AU POINT DU LOGICIEL PRIMERDEG 0.1.....		124
I.	PROTOCOLE GENERAL DU FONCTIONNEMENT DE PRIMERDEG 0.1	125
1.1.	<i>Récupération des séquences</i>	125
1.2.	<i>Détermination de la position et découpage des séquences</i>	125
1.3.	<i>Alignements multiples des séquences nucléiques</i>	127
1.4.	<i>Calcul du consensus</i>	127
II.	PRESENTATION ET UTILISATION DU LOGICIEL	130
2.1.	<i>Présentation de PrimerDeg 0.1</i>	130
2.2.	<i>Utilisation du logiciel</i>	131
2.2.1.	Utilisation de l'onglet ProtRegion2DNA.....	131
2.2.2.	Utilisation de l'onglet PrimerDeg.....	133
	<i>Conclusion et Perspectives</i>	134
CHAPITRE 9 IDENTIFICATION D'UN NOUVEAU LIPOPEPTIDE NON RIBOSOMIAL CHEZ <i>BACILLUS THURINGIENSIS</i> SEROVAR <i>BERLINER</i> : LA KURSTAKINE.....		136
I.	ANALYSES BIOLOGIQUES ET BIOCHIMIQUES	136
1.1.	<i>Recherche de lipopeptides</i>	137
1.1.1.	A partir de surnageant de culture.....	137
1.1.2.	Analyses moléculaires.....	141
II.	ANALYSES BIOINFORMATIQUES DES PROTEINES DETECTEES PAR LES AMORCES SPECIFIQUES DE LIPOPEPTIDES	145
	<i>Conclusions et Perspectives</i>	148
DISCUSSION GENERALE.....		150
CONCLUSIONS GENERALES.....		153
RÉFÉRENCES BIBLIOGRAPHIQUES.....		154
ANNEXES		170

Liste des figures

Figure 1. Structures de quelques peptides non ribosomiaux : A-E représentant les 5 catégories de peptides en fonction des diverses activités biologiques (Schwarzer et al. 2003).

Figure 2. Activation d'un acide aminé en aminoacyl adénylate par le domaine d'adénylation. La réaction nécessite une molécule d'ATP et du Mg^{2+} et provoque la libération d'un pyrophosphate (Sieber and Marahiel 2005).

Figure 3. Thiolation de l' aminoacyl adénylate activé (Sieber and Marahiel 2005)

Figure 4. Formation d'un pont peptidique par le domaine C. Donneur putatif (d) et accepteur (a) (Sieber and Marahiel 2005)

Figure 5. Activité d'un domaine de méthylation (Schwarzer et al. 2003)

Figure 6. Activités des domaines de formylation (A) et d'oxydation (B) (Schwarzer et al. 2003)

Figure 7. Structure primaire de l'opéron de la bacitracine synthétase (résidus d'acides incorporés par les synthétases BA1, BA2 et BA3)

Figure 8. Schéma de la bacitracine A (Schwarzer et al. 2003)

Figure 9. Biosynthèse non-linéaire de la vibriobactine (a. DHBA, b. thréonine, C. norspermidine)

Figure 10. Schéma de la vibriobactine (Mootz et al. 2002)

Figure 11. Biosynthèse itérative de l'entérobactine (Mootz et al. 2002)

Figure 12. Principaux domaines impliqués dans la synthèse de la déoxyérythronolide (précurseur de l'érythromycine). ACP: Acyl carrier protein, AT: Acyl transférase, DH: Déshydratase, ER: Enoyl réductase, KR: Kétoréductase, KS: Kétosynthase, TE: Thioestérase.

Figure 13. Modifications Post-PKS de la synthèse de l'érythromycine A

Figure 14. Représentation schématique des opérons (ORFs, domaines NRPSs ou PKSs et acides aminés incorporés par les différents modules encodés par la machinerie catalytique responsable de la biosynthèse de membres représentatifs de chaque famille de lipopeptides produit par *Bacillus subtilis*. A: domaine adénylation, ACP (Acyl carrier protein), AL: Acyl-CoA ligase, AMT: amino transférase, C: domaine condensation, E: domaine épimerisation, KS: kétéosynthase, MCT: Malonyl-CoA transacylase, T: domaine thiolation, Te: domaine thioestérase.

Figure 15. Opéron de la surfactine synthétase (Sieber and Marahiel 2005)

Figure 16. Carte génétique de pJET1/Blunt (Source manuel Fermentas)

Figure 17. Carte génétique de pGEMT-Easy (Source manuel Promega)

Figure 18. Photographie de soubala

Figure 19. Organisation des modules et domaines des synthétases non ribosomiales de *Lactobacillus plantarum* WCSF1, acides aminés activés par les gènes npsA et npsB, X : Substrat activé non reconnu; Y : module non reconnu comme étant de type non ribosomal, ACol : monomère non identifié

Figure 20. Représentation schématique des alignements de séquences nucléiques et des motifs conservés obtenus par le logiciel Vector NTI (A : motif ayant servi à définir A7.d1, amorce sens, B : motif ayant servi à définir A10.d1, amorce non sens), les séquences des amorces étant dans le sens 5' 3'

Figure 21. Détection sur gel d'agarose des fragments amplifiés par PCR sur les ADNs des souches témoin positif (Gel A: T0 témoin amorces sans ADN, Puits 1 : ADN *L. plantarum* WCSF1 avec amorce A7.1/T1.1, Puits 2 : ADN *L. plantarum* 8826 avec amorce A7.1/T1.1, Puits 3 : ADN *L. plantarum* WCSF1 avec amorce A10.4/C3.4, Puits 4 : ADN *L. plantarum* 8826 avec amorce A10.4/C3.4) (Gel B : Puits 1 : ADN *L. plantarum* WCSF1 avec amorce A7.d1/A10.d1, Puits 2 : ADN *L. plantarum* 8826 avec amorce A7.d1/A10.d1) (Gel C. Puits 1 : témoin sans ADN, Puits 2 : DNA *L. sakei-sakei* 16 avec les amorces Alb1-F/Tlb1-R, Puits 3 : DNA WCSF1 avec les amorces Alb1-F/Tlb1-R) M : Marqueur de Taille O' Gène Ruler

Figure 22. Séquences des fragments amplifiés par les amorces dégénérées (A. Fragment 750 bp de *L. plantarum* WCSF1 avec les amorces Alb1-F/Tlb1-R, B. Fragment 450 bp de *L. sakei-sakei* 16 avec les amorces Alb1-F/Tlb1-R, C. Fragment 300 bp amplifié par les amorces A7.d1/A10.d1 chez *L. sakei-sakei* 16)

Figure 23. Détection sur gel d'agarose des fragments amplifiés par PCR effectués sur les ADNs des souches 16, 65, 110 et 111 avec les amorces Alb1-F/Tlb1-R (Témoin: *L. plantarum* WCSF1)

Figure 24. Séquences du fragment de 1080 bp amplifié par les amorces dégénérées Alb1-F/Tlb1-R de la souche Lb111

Figure 25. Représentation schématique des alignements de séquences nucléiques et des motifs conservés (encadrés en bleu) qui ont servi à la définition des amorces spécifiques Am1-F et Tm1-R

Figure 26. Résultats des réactions PCR in vitro sur l'ADN des souches de *Bacillus subtilis* en notre possession (M: O'Gene ruler, T°: échantillon sans ADN, T1: ADN avec amorces PrBac, 1: ADN *B. subtilis* 168, 2: ADN *B. subtilis* ATCC 21332, 3: ADN *B. subtilis* ATCC 6633, 4: ADN *B. subtilis* ATCC 9943)

Figure 27. Photographie de la souche **0S6B021**

Figure 28. Résultats PCR sur gel d'agarose de l'ADN de souches testées avec les amorces B-K1-F/ B-K1-R1 (M = Trak-Ilt 100 DNA Ladder, T- : témoin négatif avec DNA sans amorces, les puits 18, 19, 20, 21 et 22 sont des souches d'intérêt testées), le puits 23 une souche témoin de *Bacillus subtilis*.

Figure 29. Séquence du fragment de 1100 bp amplifié par les amorces BK1-F/BK1-R1 de la souche S 13

Figure 30. Séquence du fragment de 1100 bp amplifié par les amorces BK1-F/BK1-R1 de la souche S 21

Figure 31. Détermination de la tension de surface des souches issues des condiments fermentés

Figure 32. Spectres obtenus à partir de la souche S 6 (A. colonie sur milieu solide GN, B. surnagent sur BN)

Figure 33. Séquence des fragments PCR (A. souche 6 avec As1-F/Ts2-R, B. souche 21 avec As1-F/Ts2-R, C. Souche 21 avec Am1-F/Tm1-R)

Figure 34. Protocole général de fonctionnement du logiciel PrimerDeg 0.1 (en chiffre rouge, les différentes étapes critiques mises en œuvre pour définir des amorces)

Figure 35. A. Un extrait de la séquence protéique de la surfactine synthétase A (en bleu : le domaine d'adénylation débute à la position 478 et prend fin à l'acide aminé en position 956, en rouge : le domaine de thiolation débute à la position 978 et prend fin à l'acide-amino en position 1040)

Figure 36. Alignement multiple de séquences nucléiques

Figure 37. Interface montrant les deux onglets du logiciel PrimerDeg 0.1.

Figure 38. Exemple de découpage en domaines par NCBI (protéine SrfA-A)

Figure 39. Interface de l'onglet ProtRegion2DNA du logiciel PrimerDeg 0.1

Figure 40. Interface de l'onglet PrimerDeg du logiciel PrimerDeg 0.1

Figure 41. Photographie des tests antifongiques BT S 45 et BT 1915

Figure 42. Détection de la présence d'iturines dans des surnageants de culture des souches *B. thuringiensis* serovar berliner 1915 et S 45 (A. Témoin standard iturine A concentré à 500 mg.mL⁻¹; B. souche 1915; C. souche S 45)

Figure 43. Spectres obtenus sur milieux DSM solide (A. *B. thuringiensis* S 45 ; B. *B. thuringiensis* 1915) et liquide (C. *B. thuringiensis* S 45 ; D. *B. thuringiensis* berliner 1915)

Figure 44. Séquences des fragments amplifiés par les amorces Abl1-F/Tbl1-R (A. ADN S 45, B. ADN 1915, C. ADN 6633, D. ADN 168)

Figure 45. Proposition d'un schéma de l'opéron responsable de la molécule détectée chez *B. thuringiensis* serovar berliner à partir des gènes de la souche *B. cereus* G9842

Figure 46. Structure de la kurstakine de *B. thuringiensis* kurstaki HD-1 (Hathout et al. 2000).

Liste des Tableaux

Tableau 1. *Motifs protéiques conservés des domaines NRPS*

Tableau 2. *Structure primaire de quelques lipopeptides de Bacillus*

Tableau 3. *Caractéristiques des souches de Lactobacillus étudiées*

Tableau 4. *Coefficients affectés à chaque acide nucléique dégénéré*

Tableau 5. *Conditions expérimentales de programmation de l'appareil PCR pour l'utilisation des amorces*

Tableau 6. *Caractéristiques des amorces utilisées pour l'identification des souches isolées*

Tableau 7. *Caractéristiques des amorces spécifiques NRPS Lactobacillus (Vector NTI et AmplifX 1.4.4)*

Tableau 8. *Taille des fragments obtenus par PCR avec les 4 couples d'amorces définies sur l'ADN de 5 souches de lactobacilles*

Tableau 9. *Pourcentage de similarité entre les domaines d'adénylation de NpsA et des protéines de prophage détectées par les amorces A7.d1/A10.d1 (Vector NTI)*

Tableau 10. *Activités antifongiques de bactéries lactiques isolées de Clupea harengus fumé*

Tableau 11. *Analyse comparative des synthétases de lipopeptides NRPS de souches de Bacillus par Needle*

Tableau 12. *Potentiel de synthèse de lipopeptides non ribosomiaux et présence de gènes correspondants mis en évidence chez les souches de Bacillus subtilis*

Tableau 13. *Caractéristiques des amorces spécifiques pour la détection des synthétases de lipopeptides de Bacillus*

Tableau 14. *Tests PCR in silico réalisés avec les amorces lipopeptides de Bacillus sur les différents ORFs*

Tableau 15. *Résultats de Blast des séquences de fragments obtenus par PCR avec les amorces dégénérées*

Tableau 16. *Détermination des pourcentages d'accrochage des amorces Af2-F/Tf1-R sur les gènes de la fengycine, la surfactine et de la plipastatine*

Tableau 17. *Souches isolées sur BN (à 30°C et 37°C) à partir du Soumbala*

Tableau 18. *Souches isolées sur BN (à 30°C et 37°C) à partir du Bikalga (B)*

Tableau 19. *Souches isolées du Bikalga (B) et du Soumbala (S) sur le milieu MRS (30°C et 37°C)*

Tableau 20. *Distribution du nombre de souches isolées en fonction de la température et des aliments*

Tableau 21. *Activités antifongiques et antimicrobiennes des souches isolées de condiments africains*

Tableau 22. *Caractéristiques morphologiques des 22 colonies de souches sélectionnées*

Tableau 23. *Divers caractéristiques des souches d'intérêt*

Tableau 24. *Profil génotypique pour l'identification moléculaire des souches criblées*

Tableau 25. *Taille des bandes amplifiées par couple d'amorce*

Tableau 26. *Résultats de Blast des séquences de fragments obtenus par PCR avec les amorces dégénérées*

Tableau 27. *Répartition du nombre de nucléotides par colonne*

Tableau 28. *Exemple de la détermination des FDN*

Tableau 29. *Activités antifongiques des souches de Bacillus thuringiensis*

Tableau 30. *Masses des pics détectés sur colonie (C) et dans les surnageants de culture (S) de souches B. thuringiensis S 45 et 1915 obtenus par Maldi-TOF*

Tableau 31. *Analyse comparative des synthétases de la famille des lipopeptides ituriniques*

Tableau 32. *Caractéristiques des amorces spécifiques de bacillomycine déterminées à l'aide du logiciel AmplifX 1.4.4 (conditions de match par défaut)*

Tableau 33. Résultats PCR in vitro par les amorces AmI-F/TmI-R et AblI-F/TblI-R (- pas de bandes, + présence de bandes)

Tableau 34. Résultats Blast des amplicons obtenus par PCR avec les amorces AblI-F et TblI-R

Tableau 35. Signatures protéiques des sites actifs de la protéine codant l'acide aminé en position 5 de la kurstakine (protéine cible)

Liste des abréviations

aa : acide aminé
Å : Angström
ACP : Acyl carrier protein
Amp : Ampicilline
ATP : Adénosine-5'-triphosphate
ARN : Acide Ribonucléique
ADN : Acide Désoxyribonucléique
BLAST: (Basic Local Alignment Tools)
bp : base paire
cm : centimètre
CDS : CoDing Sequence
CoA : coenzyme A
CMC : Concentration Micellaire Critique
Da: Dalton
kDa: Kilodalton
DEBS: désoxyérythronolide-B-synthase
DHB: dihydroxybenzoate
DNTP: 2'-désoxynucléoside-5'-triphosphate
EDTA: Acide Ethylène Diamino Tétra Acétique
FAS: Fatty Acid Synthase
h: heure
HIV: Human Immunodeficiency Virus
HPLC: High performance liquid chromatography
IPTG: Isopropyl-β-D-thiogalactoside
IUPAC: International Union of Pure and Applied Chemistry
LC/MS: Liquid chromatography/mass spectrometry
L: litre
Maldi-TOF: Matrix assisted laser desorption ionization-time of flight
Min : minutes
 mN.m^{-1} : milliNewtown par mètre
NRPS: Non Ribosomal Peptide Synthetase
ORF: Open Reading Frames
PCP: Peptidyl carrier protein ou domaine de thiolation
PCR: Polymerase chain reaction
PKS : Polykétide synthase
ppan : 4'- phosphopantéthéine
 PP_i : pyrophosphate inorganique
SA : Acide Salicylique
s : seconde
TFA : Acide Trifluoroacétique
v/v : Volume par volume
w/v : Poids par volume
§ : Paragraphe

Résumé

Les synthétases non ribosomiales sont des complexes multienzymatiques qui synthétisent des peptides atypiques par un mécanisme de thiotemplate indépendant des ribosomes. Ces peptides possèdent un large spectre d'activités biologiques et sont utilisés dans les domaines de la médecine et de la recherche biologique. Leur application dans d'autres domaines comme l'agriculture et l'environnement est de plus en plus envisagée à cause de leurs fonctions diverses et surtout pour leur innocuité par rapport aux analogues chimiques. Plusieurs microorganismes posséderaient la capacité de synthétiser ces molécules particulières.

Nos travaux se sont concentrés sur la mise au point d'une nouvelle approche moléculaire de criblage par la réaction de polymérisation en chaîne, dans le but de détecter des souches de *Lactobacillus* et de *Bacillus* ayant le potentiel de synthèse de ces peptides non ribosomiaux. Ainsi, plusieurs couples d'amorces dégénérées spécifiques (As1-F/Ts2-R, Am1-F/Tm1-R, Af2-F/Tf1-R, Ap1-F/Tp1-R et Abl1-F/Tbl1-R) ont été définis à partir des gènes connus de synthétases impliquées dans la biosynthèse de ces molécules. Ces amorces ont permis de révéler le potentiel de synthèse de molécules non ribosomiales des souches de *Bacillus cereus* LMG 2098 et de *Bacillus thuringiensis serovar berliner* 1915, puis ont confirmé la présence de gènes NRPS dans d'autres bacilles déjà connus pour la production de ces molécules.

La combinaison de ce criblage moléculaire aux analyses biochimiques classiques et aux outils bioinformatiques a permis la découverte d'une nouvelle famille de lipopeptides non ribosomiaux : les kurstakines.

PARTIE I.

ETUDES BIBLIOGRAPHIQUES

Introduction générale

Certains champignons et bactéries produisent des molécules peptidiques jouant un rôle primordial dans le développement et la survie de la cellule qui sont synthétisées par une voie différente de la voie ribosomiale classique.

Cette voie de synthèse substitue aux ribosomes, des complexes multienzymatiques pour la synthèse protéique et n'utilise donc pas la traduction des molécules d'ARN. Ce type d'enzyme est appelé **Non Ribosomal Peptide Synthetase (NRPS)**, selon l'appellation anglo-saxonne. L'intérêt de cette voie de synthèse réside dans le fait qu'elle produit une diversité de molécules décrites comme ayant un grand nombre d'activités biologiques.

Toutes les molécules de type non ribosomal décrites à ce jour sont synthétisées par des microorganismes. En effet, leur synthèse par des organismes supérieurs n'a pas encore été démontrée, on ignore toutefois si, certains cyclotides comme les circulines (un peptide macrocyclique anti-HIV découvert chez *Chassalia parvifolia*, une plante d'origine africaine), de nature structurale proche des molécules non ribosomiales, sont synthétisés par cette voie (Daly *et al.* 1999; Koltay *et al.* 2004).

Pendant ces deux dernières décennies, les études sur les NRPS se sont développées. Le nombre de publications notamment sur les activités biologiques des molécules produites et sur l'analyse de la structuration de ces NRPS s'est considérablement accru. Le mode de synthèse de ces molécules, mais aussi leurs fonctions, font d'elles des métabolites dits secondaires qui n'interviennent donc pas directement dans la croissance cellulaire.

Ces molécules peptidiques peuvent présenter des activités antifongiques, antivirales, immuno-modulatrices, toxiques ou cytostatiques pouvant ainsi bloquer la synthèse cellulaire. On y trouve aussi des surfactants, qui améliorent la miscibilité entre les phases aqueuses et organiques, potentiel recherché par les industries cosmétique et agroalimentaire, ainsi que des sidérophores qui permettent de capter dans l'environnement extérieur le fer nécessaire à la croissance de certains microorganismes.

De nombreuses souches appartenant notamment aux genres *Bacillus*, *Pseudomonas*, *Streptomyces* ont la capacité de produire ces molécules NRPS, ce qui accroît leur importance applicative dans les domaines de l'agriculture, de l'environnement, de la médecine. Peu d'études sur ces molécules ont permis jusqu'à présent de mettre au point des outils capables de détecter les souches susceptibles de les synthétiser car la plupart des recherches récentes sur les NRPS se sont davantage concentrées sur la structure de ces protéines afin de déterminer leur

fonction, sur l'organisation et le fonctionnement des enzymes impliquées dans la synthèse de ces molécules.

Pour cette étude, nous nous sommes intéressés à deux genres bactériens, le genre *Lactobacillus* qui appartient aux bactéries lactiques et le genre *Bacillus*.

Ces deux genres appartiennent aux groupes des bactéries dites GRAS (Generally Recognized As Safe) car elles ont l'avantage d'être non pathogènes et sont inoffensives pour l'homme. Leur innocuité réside essentiellement dans le fait qu'elles sont fondamentalement dérivées des bactéries commensales de la flore muqueuse normale ou sont intensivement utilisées dans la nourriture industrielle (Adams and Marteau 1995).

Les bactéries lactiques sont utilisées depuis des siècles pour fermenter les aliments afin de mieux les conserver et d'améliorer leurs qualités organoleptiques. Certaines souches ont même été envisagées pour le traitement de maladies humaines. En effet, l'intérêt des bactéries lactiques en médecine a évolué depuis les observations de Döderlein sur le rôle des bactéries vaginales dans la prévention et l'inhibition de la croissance des germes pathogènes, par la production d'acide lactique (Holzapfel *et al.* 2001). Le développement de nouvelles technologies de biologie cellulaire a ouvert une perspective d'utilisation de ces bactéries comme agents biothérapeutiques. Ces souches produiraient des enzymes (lipase, lactase, estérase), des médiateurs chimiques (hormones et interleukines) et des molécules pouvant stimuler une réponse immunitaire locale. En effet, des souches recombinantes de *Lactococcus lactis subsp. cremoris* ont été envisagées pour produire de façon importante de la riboflavine (vitamine B2). Intégrées dans des aliments fermentés, elles entraîneraient une hyperproduction *in situ* de cette vitamine rendant ainsi inutile toute supplémentation. Ces souches trouveraient leur indication dans l'aribo flavinose, le paludisme, la migraine et dans l'acidose lactique (acidité trop élevée du sang due à une accumulation importante d'acide lactique) induite par les analogues nucléosidiques (Burgess *et al.* 2004).

Malgré leur intérêt croissant pour les industries agro-alimentaires, la nutrition et la médecine, de rares cas d'infections de certaines d'entre elles, plus ou moins sévères, ont été rapportées chez l'homme. En effet, le genre *Leuconostoc* serait responsable respectivement de bactériémie et de méningite chez les adultes et les enfants alors que certains lactobacilles sont impliqués dans des endocardites (Aguirre and Collins 1993). Tout récemment, Mofredj *et al.* (2007) ont rapporté la pathogénicité de *Lactococcus garvieae* dans trois cas d'endocardite sur prothèse et deux cas de bactériémie dont une endocardite native. Ils conclurent néanmoins que la rareté des cas et leur pronostic favorable ne devraient pas remettre en cause l'intérêt industriel et médical de ces bactéries.

En 2003, Kleerebezem *et al.* ont pour la première fois mis en évidence la présence de gènes de type NRPS chez *Lactobacillus plantarum* WCSF1. Il s'est alors avéré utile, dans un premier temps, de vérifier si les collections de lactobacilles en notre possession étaient capables de produire ces molécules particulières.

Dans un second temps, une collection de bacilles a été criblée pour la recherche de nouvelles molécules NRPS. En effet, certaines souches de bacilles (*B. subtilis*, *coagulans*, *amyloliquefaciens*, *pumilus*, *cereus*, *thuringiensis*, *brevis* et *licheniformis*) sont décrites comme produisant une gamme de molécules non ribosomiales qui leur confère un pouvoir de biopreservation dans la lutte biologique c'est-à-dire qu'elles sont douées d'attitude antagoniste vis-à-vis de pathogènes comme des moisissures et certaines levures. De plus en plus d'études corrélient l'activité de ces souches à leur capacité à produire des molécules non ribosomiales.

L'objectif principal de cette étude est donc de mettre au point des outils moléculaires capables de détecter rapidement des bactéries des genres *Lactobacillus* et *Bacillus* pouvant synthétiser des molécules non ribosomiales et en particulier des lipopeptides et ce, afin de mettre en évidence de nouvelles molécules.

Après une revue succincte des études concernant les peptides non ribosomiaux, nous présenterons, l'ensemble des travaux que nous avons réalisés pour la détection et l'identification des gènes de synthétases impliquées dans la synthèse des molécules non ribosomiales ainsi qu'un outil bioinformatique mis au point pour l'automatisation des procédures.

Chapitre 1 Les peptides non ribosomiaux

I. Historique des peptides non ribosomiaux

La pénicilline, le plus célèbre des antibiotiques et la cyclosporine, le plus utilisé des immunosuppresseurs, ont en commun leur appartenance à la famille des peptides non ribosomiaux. À la fin des années cinquante (soit trente ans après la découverte de la pénicilline), personne ne pouvait imaginer que le mécanisme mis en œuvre pour la synthèse de ces peptides s'affranchisse des règles formulées au cours de la décennie précédente notamment par le biologiste américain Joshua Lederberg, Prix Nobel en 1958. Pourtant, la structure inhabituelle de la plupart des antibiotiques peptidiques suggère fortement qu'ils ne peuvent être synthétisés par la voie ribosomiale :

- Ces peptides renferment une très forte proportion d'acides aminés ne pouvant être reconnus par des ribosomes notamment ceux de conformation D.
- Ces peptides arborent fréquemment des structures cycliques impliquant parfois des liaisons ester que l'on ne trouve jamais dans les peptides synthétisés par les ribosomes sauf après une modification post-traductionnelle.

C'est au début des années soixante que des chercheurs japonais et norvégiens qui travaillent sur la biosynthèse d'un antibiotique peptidique, la gramicidine S observent qu'une fois traitée avec une enzyme détruisant l'ARN, ou en présence d'une substance abolissant toute synthèse ribosomique, les extraits cellulaires de la bactérie sont toujours capables de synthétiser cette molécule (Tomino *et al.* 1967). Ils concluent sur l'existence d'un mécanisme de synthèse peptidique indépendant de l'ARN et découvrent ensuite chez la bactérie concernée de grandes enzymes dont la fonction demeure inconnue. Seraient-elles impliquées dans la synthèse de la gramicidine ?

C'est en 1971 que Fritz Lipmann et ses collègues de l'Université Rockefeller, à New York, en isolant alors deux des longues enzymes suspectées d'intervenir dans la formation de l'antibiotique, démontrent qu'elles sont bien impliquées dans la synthèse du peptide.

Ces enzymes sont généralement de grandes tailles. La cyclosporine synthétase est de nos jours, la plus longue chaîne polypeptidique naturelle connue : elle est composée de plus de 15 000 acides aminés alors que le peptide cyclique qu'elle synthétise, la cyclosporine, n'est composé que de 11 acides aminés (Lawen and Zocher 1990; Hoppert *et al.* 2001).

La synthèse non ribosomiale permet donc à la cellule de s'affranchir des contraintes de sélection des acides aminés inhérentes à la synthèse ribosomiale, et de produire des peptides aux

particularités structurales sans égal. Or, ce sont ces particularités structurales qui, le plus souvent, confèrent la singulière activité biologique aux antibiotiques et aux autres peptides produits par les synthétases. Une activité biologique qui, pour nombre de ces peptides, demeure encore inconnue. Des études de plus en plus ciblées et nombreuses sur les peptides issus de cette voie permettent d'accroître la compréhension de ce mécanisme et les connaissances sur leurs activités biologiques particulières.

II. Diversités et activités biologiques des Peptides Non Ribosomiaux

On regroupe sous le vocable « Peptides Non Ribosomiaux ou NRP » l'ensemble des molécules peptidiques issues d'une synthèse non ribosomiale. Ces molécules sont soit issues exclusivement du mécanisme NRPS soit proviennent de la combinaison du mécanisme NRPS avec d'autres mécanismes comme le système Polykétide Synthase ou PKS.

Les peptides non ribosomiaux sont souvent regroupés selon leur activité biologique, selon l'homologie de la structure de la synthétase impliquée dans leur synthèse et quelques fois selon la nature des molécules.

L'activité biologique des peptides non ribosomiaux est fortement liée à leur structure tertiaire (Peypoux *et al.* 1999). Ainsi leur diversité structurale (linéaire, branchée, cyclique ou autre) leur confère un large spectre d'activités que l'on classe empiriquement en 5 catégories.

- Les antibiotiques ou les précurseurs d'antibiotiques

Les molécules telles que la gramicidine S (Byford *et al.* 1997) et la vancomycine (Mootz and Marahiel 1997) possèdent une activité antibiotique alors que l'ACV-tripeptide (Hori *et al.* 1989) est un précurseur d'antibiotiques tels que la pénicilline et la céphalosporine.

La plupart de ces antibiotiques sont produits par le genre *Bacillus* et le groupe des actinomycètes. Les produits les plus connus sont la gramicidine et la tyrocidine produits par *Bacillus brevis* ou encore la bacitracine produite par *Bacillus licheniformis* et *B. subtilis* (Podlesek and Grabnar 1987). Les actinomycètes sont plutôt connus pour la production d'antibiotiques glycopeptidiques comme la tubéractinomycine du genre *Streptomyces* utilisée contre la tuberculose.

- Les agents immunosuppresseurs

Le plus connu est certainement la cyclosporine A (Lawen and Zucher 1990) efficace après la transplantation des organes contre les phénomènes de rejets. Elle est aussi utilisée dans le traitement des polyarthrites.

- Les agents cytostatiques

Ce sont, en général, des molécules qui pourraient être utilisées dans le traitement des cancers comme les épothilones (Nettles *et al.* 2004) et la bléomycine A2 (Brahim *et al.* 2008). Les épothilones sont des macrolactones produites par une myxobactérie *Sorangium cellulosum* qui sont, actuellement, testés en phase clinique.

- Les sidérophores

Ce sont des substances capables de chélater le fer et sont généralement produites lors d'un déficit en fer dans l'environnement du microorganisme. Une fois excrétées à l'extérieur, ces molécules fixent le fer ferrique ambiant et sont récupérées par le biais de récepteurs membranaires spécifiques.

Les sidérophores sont sans doute la 2^{ème} classe de molécules non ribosomiales, en nombre, après les antibiotiques connues à ce jour. On les regroupe principalement en 3 groupes selon la nature du groupement chélateur (catécholate, hydroxamate et hydroxy-carboxylate) de l'ion ferrique. Les sidérophores peuvent être utilisés à des fins thérapeutiques en cas de surcharge en fer ou même en cas de déficit.

- Autres activités

Les peptides non ribosomiaux peuvent être des toxines comme la victorine (Nikolskaya *et al.* 1995), des pigments comme l'indigodine (Schwarzer *et al.* 2003) et aussi possèdent des fonctions non encore déterminées comme celle de l'anabaenopeptilide 90-A.

Ces différentes molécules peuvent également être classées en fonction de leurs traits structuraux. On retrouve 4 classes majeures:

- des peptides

Les peptides sont essentiellement composés d'acides aminés comme la bacitracine (Konz *et al.* 1997).

Cette classe de molécule représente la plus importante des antibiotiques de type non ribosomal recensée à ce jour.

- des lipopeptides

Ce sont des peptides liés à des acides gras par une liaison ester ou amide. La partie peptidique est synthétisée par le mécanisme NRPS et la chaîne d'acide gras provient du pool intracellulaire et est ou non modifiée par une polykétide synthase.

Les plus étudiés sont ceux produits par le genre *Bacillus* et qui jouent essentiellement un rôle de biosurfactant comme la surfactine (Deleu *et al.* 2003). En effet, la plupart des lipopeptides, malgré d'autres potentiels biologiques, sont d'abord étudiés pour leur propriété de biosurfactants. Cette classe particulière de molécules fait l'objet d'une description plus approfondie dans le Chapitre 2.

- des glycopeptides

L'éremomycine (Berdnikova *et al.* 2009) et la vancomycine (Hubbard and Walsh 2003) font partie de la classe des glycopeptides non ribosomiaux. Les glycopeptides sont principalement des antibiotiques inhibant la synthèse du constituant principal de la paroi des bactéries : le peptidoglycane. Ils sont néanmoins inefficaces contre les bactéries Gram- car elles possèdent une membrane externe qui isole la paroi de l'extérieur.

- des peptaïbols

Ce sont des molécules polypeptidiques comportant 15 à 20 résidus monomériques liés à un alcool. Une autre particularité réside dans leur forte proportion en acide aminoisobutyrique (Aib). On y retrouve des molécules comme l'alaméthicine (Psurek *et al.* 2005), un antibiotique produit par *Trichoderma viride* et la zervamicine (Ovchinnikova and Murashev 2006) produit par *Emericellopsis salmosynnemata* qui est efficace contre la malaria ou le paludisme (maladie la plus mortelle au monde et qui sévit principalement dans les pays tropicaux en voie de développement). D'autres molécules telles que les efraeptines produites par *Tolypocladium niveum subsp. inflatum*, en inhibant l'activité mitochondriale des parasites notamment de *Plasmodium falciparum*, peuvent également être utilisées pour un traitement curatif chez des patients atteints de malaria (Nagaraj *et al.* 2001).

La figure 1 présente les structures de quelques peptides non ribosomiaux. Les molécules peuvent notamment être branchées comme la vibriobactine et la bacillibactine, circulaires comme l'iturine et la cyclosporine, partiellement cycliques comme la fengycine et la bacitracine, contenant un double cyclique comme l'aciculitine.

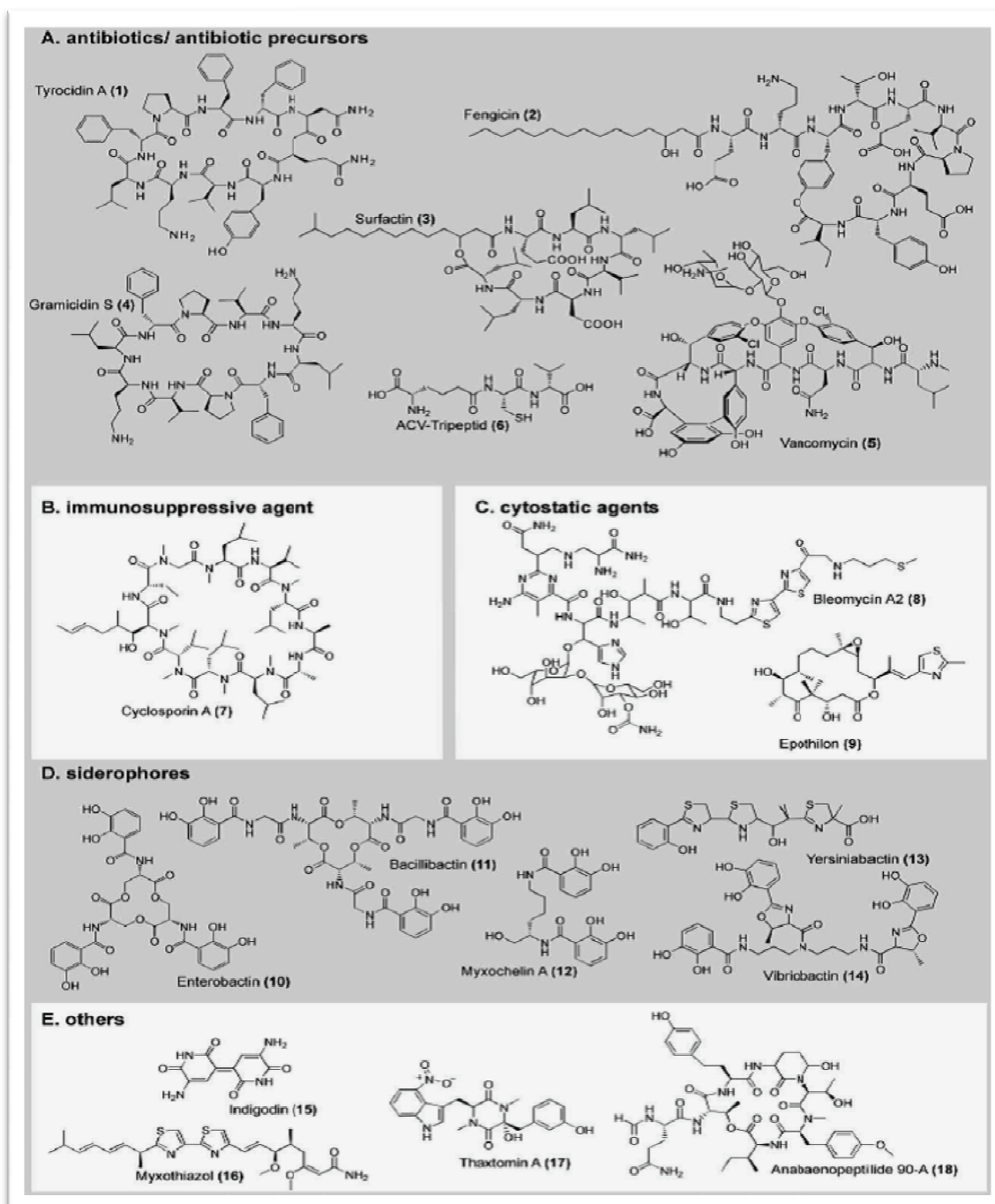


Figure 1. Structures de quelques peptides non ribosomiaux : A-E représentant les 5 catégories de peptides en fonction des diverses activités biologiques (Schwarzer *et al.* 2003).

Il existe un grand nombre de molécules non ribosomiales produites par diverses souches bactériennes et fongiques. Un portail Web (Norine) est entièrement dédié à répertorier l'ensemble des peptides non ribosomiaux issus de tous les microorganismes. Il contient actuellement un peu plus de 1070 peptides regroupés en fonction de la structure de la molécule, de son activité biologique ou de la classe de la molécule (Caboche *et al.* 2008). Un autre lien

rassemblant tous les peptaïbols d'origine NRPS informe sur la séquence et la structure de ces peptides mais aussi sur leur fonction (Whitmore and Wallace 2004).

III. Le mécanisme de la synthèse non ribosomiale

Les peptides non ribosomiaux peuvent être produits par les systèmes NRPS combinés ou non avec les PKS. Nous décrirons en détail dans un premier temps les domaines impliqués dans le mécanisme NRPS, car ils constituent le socle de notre étude et dans un second temps une comparaison des domaines NRPS/PKS.

Pour la synthèse d'une molécule donnée, plusieurs gènes sont souvent impliqués et sont regroupés sous forme d'opéron. Une synthétase NRPS est composée de plusieurs modules, dont le nombre peut varier de 2 à une vingtaine (Stachelhaus and Marahiel 1995). Chaque module est responsable de l'incorporation spécifique d'un monomère donné dans la chaîne peptidique en croissance. Il est lui-même divisé en sous-unités appelées domaines. Les monomères incorporés sont en général des acides aminés protéinogéniques, hydroxylés, N-méthylés, glycosylés ou sous forme D (Kessler *et al.* 2004).

3.1. Structure modulaire des NRPS

Trois types de modules peuvent être définis: les modules d'initiation qui comprennent au minimum un domaine d'Adénylation (A) et de Thiolation (T), les modules d'élongation qui comprennent au minimum un domaine de Condensation (C), un domaine d'Adénylation et un domaine de Thiolation et enfin le module de terminaison qui contient en plus du module d'élongation, un domaine de Thioestérase (Te). Cette structuration modulaire a permis de classer les différents domaines en domaines principaux et en domaines secondaires qui sont rencontrés dans la majorité des NRPS.

3.1.1. Domaines principaux

Les différents domaines principaux composant les modules sont: les domaines d'adénylation, de thiolation, de condensation et de thioestérase.

- Le domaine d'adénylation (A)

Le domaine A est composé d'environ 500-600 aa soit une taille variant de 1500 à 1800 nucléotides. Il joue un rôle spécifique dans la reconnaissance et l'activation des monomères incorporés dans le peptide (Lee *et al.* 2005).

Après la réaction avec une molécule d'ATP, l'acide aminé est converti en aminoacyl adenylate (molécule intermédiaire) avec libération d'une molécule de pyrophosphate. La figure 2 présente de façon schématique la réaction lors de cette activation.

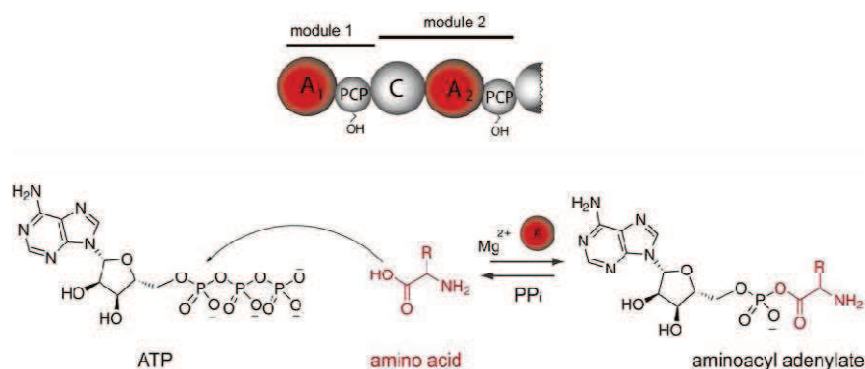


Figure 2. Activation d'un acide aminé en aminoacyl adénylate par le domaine d'adénylation. La réaction nécessite une molécule d'ATP et du Mg²⁺ et provoque la libération d'un pyrophosphate (Sieber and Marahiel 2005).

Ce mécanisme ressemble à l'activation de l'acide aminé par l'aminoacyl-tRNA-synthétase de la voie ribosomiale. L'aminoacyl-tRNA-synthétase et le domaine A appartiennent à la même famille enzymatique mais ont des structures différentes (Stryer 1996). L'analyse des domaines d'adénylation a permis de montrer qu'ils ont une taille similaire à celle de la famille des acyl-CoA synthases et des luciférases. De plus, ces enzymes catalysent la même réaction c'est-à-dire l'adénylation de substrats carboxylés, ce qui permet de les regrouper dans la superfamille des enzymes formant des « Adenylates » (Marahiel *et al.* 1997). La comparaison au niveau protéique de différents domaines NRPS entre eux a permis de déterminer des régions fortement conservées, appelées signatures des domaines ou core motifs. Ainsi, le domaine d'adénylation contient 10 régions protéiques conservées (A1-A10) (voir Tableau 1 suivant). Schwarzer *et al.* (2003) ont montré que seules les régions A4-A5 sont impliquées dans l'activation et la reconnaissance des monomères.

- Le domaine de thiolation (T)

Le domaine de thiolation (T) ou encore appelé Peptidyl Carrier Protein (PCP) est le seul domaine NRPS ne possédant pas une activité catalytique autonome. De taille comprise entre 80 et 100 aa, il est situé en aval du domaine A. Le domaine T fixe l'aminoacyl adenylate, activé par le domaine A, sur la synthétase de manière covalente par l'intermédiaire d'une liaison thioester,

avec un cofacteur phosphopantéthénique (ppan) préalablement fixé sur un résidu sérine (figure 3) (Ehmann *et al.* 2000). La transformation du domaine PCP de sa forme inactive (apo-PCP) à sa forme active (holo-PCP) est catalysée par une phosphopantéthéine transférase (protéine Sfp chez *B. subtilis*).

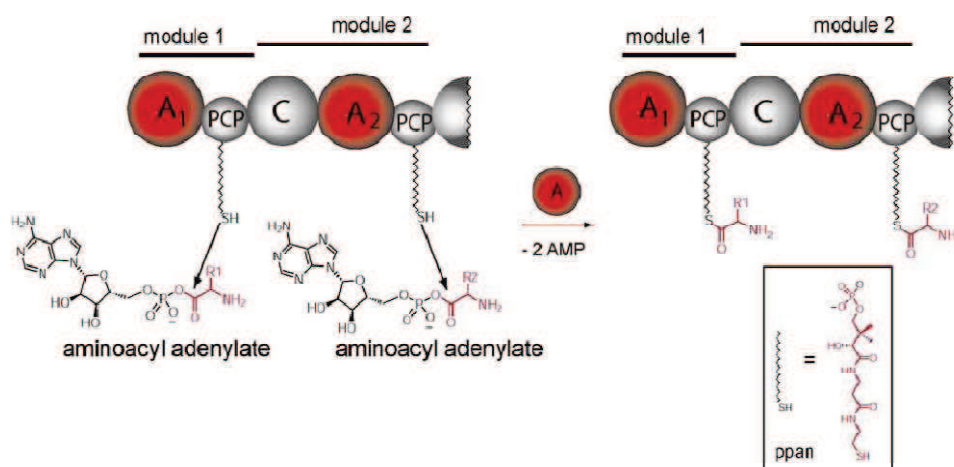


Figure 3. Thiolation de l' aminoacyl adénylate activé (Sieber and Marahiel 2005)

Le domaine T joue essentiellement un rôle de transport et d'élongation des substrats entre les domaines catalytiques. L'utilisation des domaines de transport comme le domaine T n'est pas exclusive pour les molécules NRPS. En effet, d'autres complexes multienzymatiques modulaires comme les polykétides synthases transportent les unités de malonate ou de propionate à l'aide des Acyl Carrier Protein (ACP) appartenant à la même famille que les PCP (Schwarzer *et al.* 2003). L'unique région protéique conservée des domaines de thiolation est présentée dans le tableau 1.

- Le domaine de condensation (C)

Le domaine C est composé d'environ 450 aa, et est généralement situé après les domaines A et T. Il n'est pas présent dans le premier module pour les molécules purement peptidique. Il est impliqué dans la formation d'un pont peptidique entre 2 acides aminés de modules consécutifs (Figure 4).

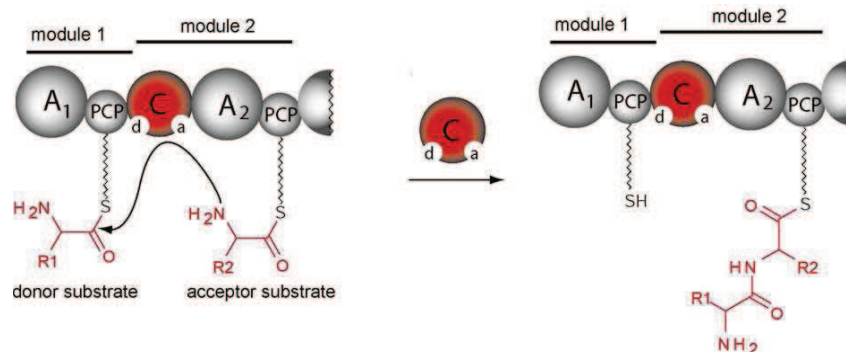


Figure 4. Formation d'un pont peptidique par le domaine C. Donneur putatif (d) et accepteur (a) (Sieber and Marahiel 2005)

Ainsi, après l'activation et la formation de la liaison covalente de l'acide aminé par le complexe A-T du module, le domaine C catalyse l'attaque nucléophile de l'acide aminé fixé sur le site PCP par le résidu fixé sur le site PCP accepteur.

Les motifs protéiques conservés des domaines de condensation ont été mis évidence pour la première fois par De Crécy-Lagard *et al.* (1995) (Tableau 1). En se basant sur ces motifs protéiques conservés et sur la phylogénie de ces domaines, certains auteurs ont montré l'existence de 3 types de domaines C. Ces domaines diffèrent par leurs fonctions spécifiques bien qu'ils soient tous impliqués dans la formation de liaisons amides entre deux monomères. Ces domaines C sont :

- le domaine C qui catalyse la formation du pont peptidique entre deux acides aminés de type L, il est appelé $^L C_L$
- le domaine C qui catalyse la formation du pont peptidique entre un résidu de type D et un autre de conformation L appelé $^D C_L$ (Clugston *et al.* 2003). Il est situé après un domaine d'épimerisation (voir § 3.1.2 suivant).
- tout récemment, Rausch *et al.* (2007) ont mis en évidence l'existence d'un troisième type de domaines C appelés « C starter ou premier C ». Ces auteurs ont montré que ces C starters sont spécifiques dans la connexion entre des acides β -hydroxylés ou carboxyliques (exemple de l'acide gras β -hydroxyl) et de l'acide L-aminé d'une chaîne peptidique d'une molécule non ribosomiale.

- **Domaine de Thioestérase (Te)**

Le domaine de thioestérase Te libère la molécule synthétisée du complexe multienzymatique. Il est composé d'environ 280 aa et se situe, de par sa fonction, à l'extrémité C-terminale du dernier module de la synthétase. La dernière étape consiste alors à régénérer la synthétase pour la production d'une nouvelle molécule pendant que la néo-molécule subit ou non des modifications

post-synthèses. Le domaine Te est également capable dans certains cas de catalyser la cyclisation du peptide et/ou son oligomérisation.

Lorsque le domaine Te est absent de la synthétase, la libération de molécules peut être effectuée par une hydrolyse acide ou par une réaction intramoléculaire.

3.1.2. Domaines secondaires

Si les domaines (C, A, T et Te) sont dits principaux car indispensables pour les étapes d'initiation, d'élongation et de termination, d'autres domaines dits « secondaires » ou « optionnels » interviennent dans la maturation ou le devenir final de la molécule y compris sa libération dans le cytoplasme. Certains modules plus complexes comportent ces sous-unités additionnelles qui catalysent des modifications supplémentaires du peptide (par exemple des épimerisations, des cyclisations, des méthylations, des oxydations ou des formylations), lesquelles contribuent significativement à la diversité structurale des monomères incorporés dans les peptides et donc à la diversité des activités biologiques des molécules NRPS.

Ces différents domaines sont :

- **Le domaine d'épimerisation (E)**

De nombreux peptides d'origine non ribosomiale contiennent des acides aminés de conformation D. Leur incorporation est effectuée par le domaine d'épimerisation (E). Il est en général composé de 450 aa et situé entre les domaines T et C.

L'épimerisation des acides L-aminés par les domaines d'épimerisation consiste au changement de la conformation du carbone C α de l'aminoacyl fixé par le domaine PCP (Pfeifer *et al.* 1995; Stachelhaus and Walsh 2000).

- **Domaine de Méthylation (M) ou (Mt)**

Le domaine de méthylation (M) est composé d'environ 420 aa et se situe en général entre les domaines T et C.

Deux types de méthylation peuvent avoir lieu notamment sur les atomes d'azote et de carbone des peptides non ribosomiaux.

Les méthylations sur les atomes d'azotes ou N-Mt concernent de façon spécifique les peptides d'origines fongiques comme la cyclosporine (Lawen and Zocher 1990) dont 7 des 11 acides aminés sont méthylés (Schwarzer *et al.* 2003).

La modification est due au cofacteur du domaine M qui est la S-Adénosyl Méthionine (SAM) dont le groupe méthyl est transféré à l'acide aminé fixé sur la synthétase.

La figure 5 présente la modification pouvant intervenir sur l'atome N lors d'une méthylation.

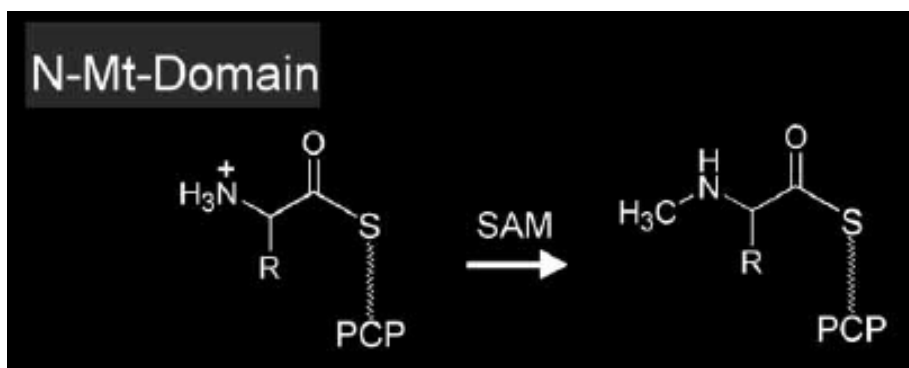


Figure 5. Activité d'un domaine de méthylation (Schwarzer *et al.* 2003)

Une autre méthylation dite C-Méthylation (C-Mt) peut subvenir sur un noyau thiazol comme dans le cas de la yersiniabactine (Miller *et al.* 2001).

- Domaines d'Oxydation et de Formylation

Le domaine de Formylation (Fo) additionne un groupement formyle à partir du tetrahydrofolate (THF) à l'extrémité NH_3 de l'acide aminé. Le radical formyl-THF fonctionne comme un coenzyme donneur de formyle (figure 6 A).

Le domaine d'Oxydation (Ox) intervient dans le processus de synthèse de peptides non ribosomiaux par la réduction d'une Flavine Mono Nucléotide en FMNH_2 et catalyse d'une double oxydation qui aboutit à la formation d'une double liaison (figure 6 B). Ces deux domaines participent ainsi à la diversification des structures peptidiques et donc de l'activité biologique.

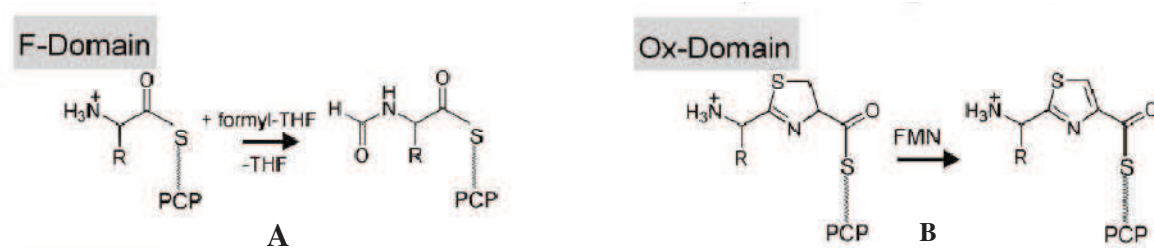


Figure 6. Activités des domaines de formylation (A) et d'oxydation (B) (Schwarzer *et al.* 2003)

- Domaines de Cyclisation (Cy) et de Réduction (R)

Les petits noyaux hétérocycliques comme les noyaux thiazoline dans la bacitracine, le noyau thiazol dans les épothilones et l'oxazoline dans les vibriobactines sont réalisés par une hétérocyclisation des chaînes d'acides aminés fonctionnels comme les cystéines, les sérines ou

les thréonines avec un carbone de la chaîne peptidique. C'est le domaine de cyclisation (Cy) mesurant environ 450 aa qui est responsable de cette hétérocyclisation. Il remplace le domaine de condensation dans certaines synthétases et dans ces cas il a une double fonction. C'est par exemple, le cas de la synthétase VibF (impliquée dans la synthèse de la vibriobactine, voir figure 9) qui contient un double domaine Cy-Cy, le premier domaine Cy intervient dans l'hétérocyclisation par la formation du noyau oxazoline entre un acide dihydroxybenzoïque et une thréonine alors que le second catalyse une condensation (Silakowski *et al.* 2000).

Les domaines de Réduction (R) sont les moins décrits et participent à réduire le substrat en modifiant sa structure initiale grâce à un coenzyme : le NADPH.

Le tableau 1 présente l'ensemble des motifs protéiques conservés dans les domaines NRPS.

Tableau 1. Motifs protéiques conservés des domaines NRPS

Domaines NRPS	Signatures	Séquence des signatures protéiques
Adénylation	A1	L(TS)YxEL
	A2 (core 1)	LKAGxAYL(VL)P(LI)D
	A3 (core 2)	LAYxxYTSG(ST)TGxPKG
	A4	FDxS
	A5	NxYGPTE
	A6 (core 3)	GELxJGx(VL)ARGYL
	A7 (core 4)	Y(RK)TGDL
	A8 (core 5)	GRxPxQVKIRGxRIELGEIE
	A9	LPxYM(IV)P
	A10	NGK(VL)DR
Thiolation	T (core 6)	DxFFxxLGGxSx
Condensation	C1	SxAQxR(LM)(WY)xL
	C2	RHExLRTxF
	C3	MHHxISDG(WV)S
	C4	YxD(FY)AVW
	C5	(IV)GxFVNT(QL)(CA)xR
	C6	HN)QD(YD)PFE
	C7	RDxSRNPL
Thioestérase	Te	GxSxG
Épimerisation	E1	PIQxWF
	E2	HHxISDG(WV)S
	E3	DxLLxAxG
	E4	EGHGRE
	E5	RTVGWFTxxYP(YV)PFE
	E6	PxxGxGYG
	E7	FNYLG(QR)
Cyclisation	Cy1	FPL(TS)xxQxAYxxGR
	Cy2	RHx(IM)L(PAL)x(ND)GxQ
	C3	D(NLI)xDxxS
	Cy4	(TS)(PA)3x(LAF)6x(IVT)LxxW
	Cy5	(GA)DFTxLxLL
	Cy6	PVVFTSxL
	Cy7	(ST)(QR)TPQVx(LI)D13xWD
Oxydation	Ox1	KYxYxSxGxxY(PG)VQ
	Ox2	GxxxG(LV)xxGxYYY(HD)P
	Ox3	IxxxYG
Méthylation	M1	VL(DE)xGxGxG
	M2	NELSxYRYxAV
	M3	VExSxARQxGxLD
Réduction	R1	V(L)(L)TG(A)TG(F)(L)GxxLL
	R2	Vx(L)(L)VR(A)
	R3	GPL(G)x(P)x(L)GL
	R4	V(Y)PYxYLxx(P)NVxxT
	R5	GYxxSKW(A)(A)E
	R6	R(P)G
	R7	YxxxxG(LF)LxxP

3.2. Biosynthèse de molécules non ribosomiales

Bien que structuré sous formes modulaires, le mode d'assemblage des substrats ou encore monomères permet 3 types de biosynthèses :

- La biosynthèse linéaire qui aboutit à la formation d'un peptide par un assemblage simple de plusieurs monomères dans l'ordre dans lequel ils sont activés par la synthétase.
- La biosynthèse itérative qui conduit à un peptide formé de plusieurs monomères et synthétisé par la répétition programmée d'un même module ou d'un ensemble de modules.
- La biosynthèse non linéaire qui assemble plusieurs monomères dans un ordre différent de celui de leur activation par la synthétase.

La multiplicité des mécanismes de biosynthèses des synthétases confère une diversité de structures supplémentaires aux peptides synthétisés et de ce fait, conduit à une large gamme d'activités biologiques. Des exemples de description de biosynthèses : linéaire (opéron bacitracine synthétase de *Bacillus*), non-linéaire : opéron vibriobactine synthétase et itérative : opéron de l'entérobactine synthétase, sont présentés dans les paragraphes suivants.

3.2.1. La biosynthèse linéaire : cas de l'opéron bacitracine chez *Bacillus*

La bacitracine est un dodécylpeptide antibiotique cyclique dont l'opéron a été mis en évidence par Konz *et al.* (1997). Cet opéron code pour 3 synthétases BA1 (598k DA), BA2 (297 kDA) et de BA3 (723 kDA) constituées respectivement de 5 modules pour BA1 et BA3 et de 2 modules pour BA2. L'ensemble des modules est responsable de l'incorporation des 12 acides aminés constituant la molécule finale. La figure 7 présente une structure primaire de l'organisation de l'opéron de la bacitracine synthétase.

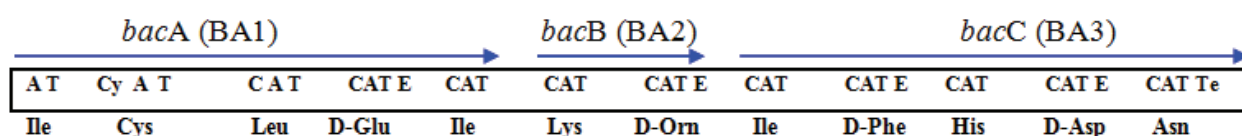


Figure 7. Structure primaire de l'opéron de la bacitracine synthétase (résidus d'acides incorporés par les synthétases BA1, BA2 et BA3)

La bacitracine synthétase fonctionne selon le type de biosynthèse linéaire. Chaque module incorpore un monomère donné dans l'ordre de l'assemblage de l'opéron pour la formation de la molécule finale. La molécule finale est obtenue par une cyclisation de la cystéine en position 2 et suivie de la formation d'un noyau thiazolique avec l'isoleucine en position 1 qui est catalysée par le produit du gène *bacA*. La figure 8 montre une représentation structurale de la bacitracine.

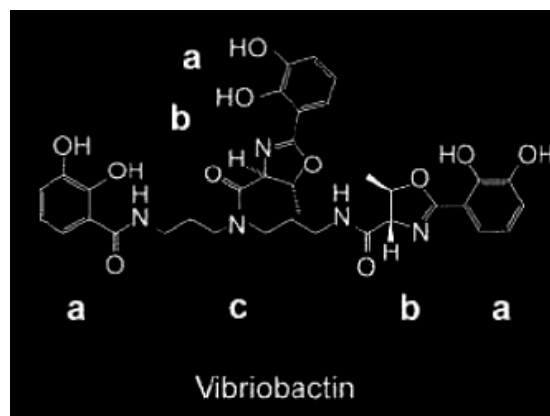


Figure 10. Schéma de la vibriobactine (Mootz *et al.* 2002)

3.2.3. Description de la biosynthèse itérative

L'itération des NRPS étant la répétition programmée de l'utilisation d'un ou plusieurs modules, elle conduit à la formation d'une molécule composée de plus d'acides aminés que le nombre de modules de la synthétase. C'est le cas par exemple de la synthétase de l'entérobactine, un sidérophore produit par *E. coli* en réponse au manque de fer dans l'environnement extérieur (Gehring *et al.* 1998). L'opéron de l'entérobactine code pour 4 synthétases (EntBDEF) groupées en deux modules permettant la reconnaissance et l'activation de deux monomères: le premier module active le 2,3-dihydroxybenzoyl et le second une L-sérine (Zhou *et al.* 2007). Trois répétitions de cet assemblage des deux monomères précités vont aboutir à la formation de la molécule finale d'entérobactine (Mootz *et al.* 2002).

La figure 11 montre le mécanisme moléculaire itératif mis en œuvre par les différents domaines dans la synthèse de l'entérobactine.

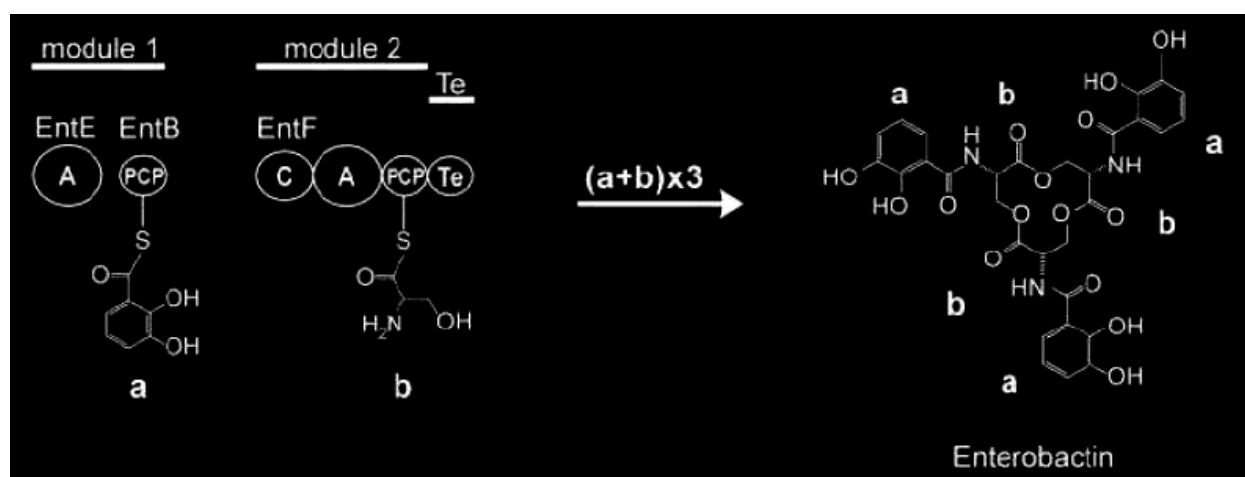


Figure 11. Biosynthèse itérative de l'entérobactine (Mootz *et al.* 2002)

La molécule finale de l'entérobactine est ainsi constituée de 6 monomères provenant de la répétition de l'action des 2 modules.

IV. Les exceptions à la synthèse non ribosomiale classique

De nombreux produits non ribosomiaux répondent à la règle dite de colinéarité qui stipule que le nombre et l'ordre des modules présents dans une synthétase déterminent la longueur et la structure du peptide résultant (Cosmina *et al.* 1993). Cependant, il existe de nombreuses exceptions telles les modes de biosynthèses non-linéaire et itérative qui aboutissent à des produits contenant des monomères qui ne sont pas organisés selon l'ordre des modules qui composent la synthétase ou encore à un nombre de monomères plus important que le nombre de modules impliqués dans leur synthèse. La structuration de certains modules ne respecte pas toujours les règles évoquées plus haut. En effet, certains modules comme celui correspondant à la synthétase SrfAD codé par le gène *srfA-D* et impliqué dans la synthèse de la surfactine ne contiennent qu'une thioestérase. On parle alors de module tronqué qui, dans ce cas est non fonctionnel. Steller *et al.* (2004) pensent néanmoins que la synthétase SrfAD pourrait intervenir dans l'activation de l'initiation de la biosynthèse de la surfactine en plus de sa fonction classique qui consiste à libérer la molécule formée du complexe multienzymatique.

L'absence de certaines séquences bordant les gènes, dénommées *linkers* et qui serviraient de guide à l'assemblage des complexes enzymatiques dans le cytoplasme (Broadhurst *et al.* 2003) peut rendre tout un opéron non fonctionnel. Enfin, la présence d'un opéron entier ne peut parfois pas suffire à synthétiser une molécule sans l'intervention de certaines protéines notamment pour l'activation du domaine T comme les phosphopantetheinyl transférase (la Sfp et la Gsp respectivement de *Bacillus subtilis* et *Bacillus brevis*). En effet, la protéine Sfp permet l'activation de la forme non active (*apo*-PCP) en sa forme active (*ppan*-PCP *holo*-form). La non activation du domaine T par la protéine Sfp chez *Bacillus subtilis* 168 est à l'origine de l'absence de synthèse de surfactine et de fengycine par cette même souche (Lee *et al.* 2005; Tsuge *et al.* 2007). Ce résultat a été confirmé par la production de la surfactine après l'ajout du gène *sfp* de *Bacillus subtilis* ATCC 21332 chez *B. subtilis* JH 642, une souche dérivée de la *Bacillus subtilis* 168 (Nakano *et al.* 1988).

Les limites des règles énoncées précédemment s'observent tout aussi bien dans la production de certaines molécules hybrides NRPS/PKS comme la yersiniabactine (Schwarzer *et al.* 2001).

V. Structure modulaire des Polykétides Synthases

Comme les NRPS, les Polykétides Synthases (PKS) sont des protéines modulaires multienzymatiques qui synthétisent des molécules actives, appelées dans ce cas des polykétides ou polycétides. Les polykétides synthases activent des petits acides carboxyliques. Les PKS sont divisés en trois grandes familles: les PKS de type I, II et III. Elles divergent entre elles principalement sur leur mécanisme de fonctionnement et dans leur organisation structurale.

Nous décrirons principalement les PKS de type I car elles sont, par leur organisation, similaires aux NRPS.

5.1. Polykétides Synthases de type I

Les PKS de type I sont constituées, comme les NRPS, de modules puis de domaines (Ginolhac *et al.* 2005). Elles produisent des métabolites complexes en utilisant un mécanisme de biosynthèse itératif ou linéaire (Gokhale *al.* 2007). Dans un mécanisme itératif, les mêmes modules sont utilisés de manière répétitive pour fabriquer le produit final. En effet les PKS de type I réalisent une succession de condensations oxydatives à partir d'acides carboxyliques dérivés de l'acide malonique. Tout comme les domaines NRPS, les PKS de type I possèdent des domaines principaux et optionnels. La figure 12 présente les différents domaines du système PKS de type I responsable de la synthèse de l'érythromycine (Staunton and Weissman 2001), une molécule antibactérienne produite par *Saccharopolyspora erythraea*. La voie de biosynthèse de l'érythromycine est composée de trois gènes PKS type I, *DEBS1*, *DEBS2* et *DEBS3*. Les trois PKS de type I résultantes utilisent comme précurseurs de la biosynthèse les acides carboxyliques comme le propionyl et le méthylmalonyl.

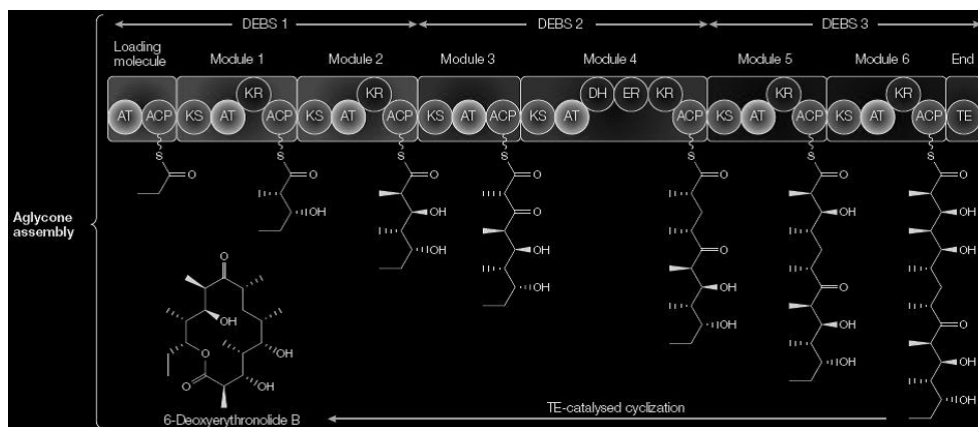


Figure 12. Principaux domaines impliqués dans la synthèse de la déoxyérythronolide (précurseur de l'érythromycine). ACP: Acyl carrier protein, AT: Acyl transférase, DH: Déhydratase, ER: Enoyl réductase, KR: Kétoréductase, KS: Kétosynthase, TE: Thioestérase.

Le domaine kéto-synthase (KS) réalise la condensation de deux substrats appartenant à deux modules adjacents sur la chaîne comme le domaine C dans le cas des NRPS. Le domaine acyl-transférase (AT) reconnaît et active le substrat comme le fait le domaine A, le domaine Acyl Carrier Protein (ACP) porte la chaîne en élongation comme le fait le domaine T dans le cas des NRPS (Hopwood 2004).

A côté de ces domaines principaux, on trouve trois domaines optionnels qui réduisent le substrat: une kéto-réductase (KR) qui réduit la fonction cétone en une fonction alcool, une déhydratase (DH) qui élimine une molécule d'eau et crée une double liaison et l'énol-réductase (ER) réduisant la double liaison. Ces domaines sont additionnels et leur action est linéaire, c'est-à-dire que le domaine DH ne peut agir que si le domaine KR a réduit la fonction cétone en fonction alcool par exemple. Leur présence ou non ne semble pas avoir d'effet sur la spécificité du domaine AT (Jenke-Kodama *et al.* 2005). Cependant, les domaines KS et KR pourraient avoir un rôle dans la stéréochimie des substrats et non dans leur nature (Staunton et Weissman 2001).

A la fin de la synthèse de la chaîne des six modules (Figure 13), le domaine thioestérase (Te) catalyse la libération et la cyclisation du noyau macrolactone du 6-déoxy-érythronolide B. Les modifications post-PKS concernent une hydroxylation, deux glycosylations et une méthylation permettant la transformation de la déoxyérythronolide en érythromycine.

La figure 13 montre la succession des modifications pour l'obtention de la molécule finale.

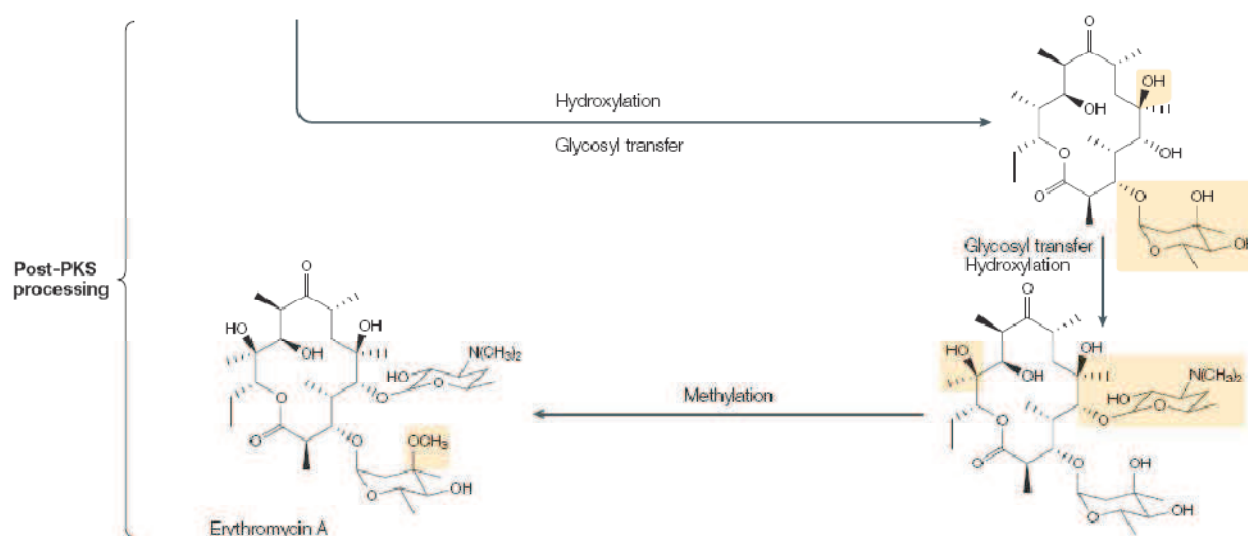


Figure 13. Modifications Post-PKS de la synthèse de l'érythromycine A

5.2. Polykétides Synthèses de type II

Les PKS de type II possèdent les mêmes domaines que les PKS de type I, mais contrairement à ces derniers qui possèdent l'ensemble de ces domaines sur un même opéron, les domaines PKS de type II sont dispersés à différents endroits du génome et les enzymes produites sont donc monofonctionnelles (Gokhale *et al.* 2007). Les PKS de ce type synthétisent principalement des polykétides aromatiques polycycliques comme l'actinorhodine (Hopwood 2004). Certaines molécules par contre synthétisées par une PKS de type II ne sont pas aromatiques comme l'entéroïcine par exemple (Moss *et al.* 2004). Les domaines minimaux d'un module de PKS de type II sont les domaines Kétosynthase Alpha et Beta ($KS\alpha$ et $KS\beta$) et l'ACP (Acyl Carrier Protein).

Le $KS\alpha$ est l'équivalent du KS et réalise la condensation des substrats sur la chaîne en élongation. Le $KS\beta$ est le domaine qui contrôle *a priori* le nombre de cycles. Tang *et al.* (2004) ont montré que ce contrôle porterait plus sur le nombre d'atomes de carbone incorporés que sur le nombre de cycles lui-même. L'ACP est particulier et semble posséder la capacité de choisir le substrat. Ceci expliquerait le fonctionnement de ces PKS en l'absence de domaine AT présent chez les PKS de type I.

5.3. Polykétides Synthèses de type III

Aussi appelées « chalcone synthase » de par leur lien avec les protéines végétales, les polykétides synthèses de type III sont des enzymes homo-dimériques itératives qui sont capables de synthétiser des polykétides, cycliques ou non, directement à partir de thioester-CoA comme le malonyl-CoA (Shen 2003). Excepté le domaine ACP, les PKS de type III possèdent les mêmes

domaines que les PKS de type II. L'ensemble des réactions enzymatiques se déroule dans un seul et même site (Staunton et Weissman 2001), signifiant l'existence d'un unique module.

5.4. Les hybrides NRPS/PKS

De nombreux exemples d'hybrides entre polykétides et peptides sont connus (Du *et al.* 2001). Dans cette étude nous nous sommes aussi intéressés uniquement aux molécules issues des mécanismes NRPS et PKS. En effet, les domaines issus des synthétases de PKS de type I et de NRPS s'associent sur un même opéron pour synthétiser une molécule dite hybride car résultant de la combinaison de deux systèmes. C'est le cas des molécules telles que la bacillaene, l'albicidine (herbicide), l'épothilone qui est un agent antitumoral (Nettles *et al.* 2004) et encore certains lipopeptides comme ceux appartenant à la famille des iturines (iturine, bacillomycine et mycosubtiline).

Le métabolisme secondaire étant générateur d'une grande diversité de composés, il n'est pas étonnant d'observer des combinaisons entre les voies principales qui synthétisent de nouvelles structures (Firn et Jones 2003). Ceci permet de produire de nouveaux composés sans toutefois créer de nouvelles voies de biosynthèses.

VI. Les différentes stratégies d'identification des molécules non ribosomiales

La complexité du mécanisme non ribosomal et la régulation des gènes qui y sont impliqués peuvent rendre difficile l'identification des molécules non ribosomiales potentiellement produites par des microorganismes.

6.1. Approches moléculaires

L'utilisation d'outils biomoléculaires pour l'identification de souches ou des molécules est de plus en plus répandue car elle permet la mise en évidence de gènes muets ou de protéines non fonctionnelles. Elle est principalement basée sur la technique dite de Polymérisation en Chaîne ou Polymerase Chain Reaction (PCR). Le principe de la PCR est d'obtenir rapidement une quantité importante et exploitable d'un segment précis d'ADN par une amplification *in vitro* de séquences nucléotidiques spécifiques. Cette technique a été inventée par Kary Mullis qui a été le premier à conceptualiser cette idée en 1983. Elle a immédiatement permis un développement révolutionnaire de la recherche des sciences du vivant. Dix ans plus tard, il recevait le prix Nobel de chimie pour l'impact de la technique PCR en biotechnologie et en médecine.

De nos jours, la PCR est l'une des techniques les plus utilisées pour la compréhension de divers mécanismes biologiques à l'échelle moléculaire. Elle comprend plusieurs variants : la Long

Range Polymerase Chain Reaction (LR-PCR), la PCR en temps réel (Q-PCR), la Reverse Transcriptase PCR (RT-PCR), etc....

Les stratégies utilisées pour l'identification de molécules de type non ribosomal ont consisté dans l'ensemble à la mise au point des conditions pour la définition des amorces, éléments clé pour le criblage de fragments spécifiques de gènes lors de l'amplification par PCR. Les amorces permettent en effet de circonscrire le fragment à amplifier en le bordant. Dans le cadre de l'amplification de gènes codant des NRPS, les auteurs ont essentiellement utilisés deux approches pour la conception de ces amorces.

La première approche consiste à définir des oligonucléotides à partir des signatures des motifs conservés protéiques (Turgay and Marahiel 1994).

Plusieurs séquences protéiques de synthétases connues et impliquées dans la synthèse de molécules non ribosomiales sont alignées puis des séquences consensus issues de ces motifs conservés ou encore « core sequences » sont déterminées. C'est ainsi que 6 core séquences contenant 3 à 8 aa ont été déterminées à partir des domaines A puis des fonctions biologiques ont été attribuées pour certains (Turgay and Marahiel 1994).

En 1992, Borchert *et al.* ont démontré qu'il était possible d'identifier et de cloner des gènes de type NRPS à partir de ces motifs conservés. Ils ont ainsi été les premiers à définir des oligonucléotides dégénérés dérivant des cores séquences protéiques KAGGA et SGTGXPKG (respectivement Core I et II) en y ajoutant les sites de restrictions pour la détection de fragments de gènes par PCR. Ces amorces leur auraient permis de montrer la présence de gènes NRPS chez *Bacillus brevis*. L'utilisation de ces amorces pour la détection de gènes de type NRPS chez *Streptomyces coelicor* par analyse moléculaire n'a pas permis leur mise en évidence soulignant ainsi, l'absence de certaines core séquences dans des génomes pourtant connus comme arborant des gènes NRPS. C'est ainsi que Turgay et Marahiel (1994) ont alors défini d'autres oligodégénérés à partir de ces signatures protéiques conservées en tenant compte de l'usage des codons au sein du genre étudié. Ils ont déterminé l'amorce sens au niveau du core II du domaine A (TGD), impliqué dans l'activation et l'hydrolyse de l'ATP et l'amorce non-sens appelée LGG à partir de la signature conservée protéique du domaine de thiolation (LGGXS) qui serait impliqué dans la formation du site de fixation d'un thioester. Ces auteurs en utilisant ces amorces (TGD/LGG) ont mis en évidence la présence de la bacitracine et de la phaséolotoxine respectivement chez *Bacillus licheniformis* et *Pseudomonas syringae*. En appliquant cette même technique, deux couples d'amorces dégénérés (KAGGA/SGTTG et TGD/LGD) ont été mis au point par Rajendran (1999) à partir des alignements des domaines d'adénylation et de thiolation de synthétases issues de *Bacillus subtilis*. Ces auteurs ont, par PCR, détecté la présence d'un

peptide putatif d'origine non ribosomiale chez *Pseudomonas fluorescens*. Ils en conclurent que si l'un des couples d'amorce ne peut amplifier un gène adéquat pour des raisons inconnues, l'autre couple en serait capable.

Ces mêmes amorces ont permis de détecter un peu plus tard divers peptides dans le genre *Bacillus* et d'autres espèces (Giacomodonato *et al.* 2001; Wu *et al.* 2005). Cette stratégie consistant à définir des amorces à partir de séquences protéiques a été également appliquée à d'autres micro-organismes tels que les mycètes filamenteux (Nikolskaya *et al.* 1995) et les cyanobactéries (Neilan *et al.* 1999; Barrios-Ilerena *et al.* 2007).

Cette stratégie de définition des amorces à partir des séquences protéiques pour la PCR (technique amplifiant des séquences nucléiques) peut s'avérer moins fiable à cause des risques d'efficacité liés au passage des séquences protéiques aux séquences nucléiques mais aussi par la présence de codons « stop » lorsqu'elle est appliquée sur d'autres microorganismes.

En outre, on ignore si les 'core sequences' ou 'motifs conservés' protéiques à partir desquels ces amorces ont été définies sont présents dans toutes les synthétases non ribosomiales. Néanmoins cette approche mise au point dans les années 90 reste de nos jours la plus utilisée pour la détection de molécules non ribosomiales par voie moléculaire.

La deuxième approche est basée sur le choix des séquences nucléiques comme matrice pour la détermination d'amorces. Cette approche aboutit à la définition d'amorces non dégénérées par l'utilisation de certains logiciels ou par la mise au point d'amorces dégénérées à partir de l'alignement de plusieurs séquences.

Certains auteurs comme Ayuso-Sacido et Genilloud (2004) ont, en alignant les séquences nucléiques des domaines d'adénylation impliquées dans la synthèse de la céphamycine (Coque *et al.* 1991), la vancomycine (Recktenwald *et al.* 2002), la balhimycine (Pelzer *et al.* 1997), l'actinomycine (Schauwecker *et al.* 1998), la pristinamycine (De Crécy-Lagard *et al.* 1997) et de la chloroérémomycine (Trauger and Walsh 2000), défini des amorces dégénérées à partir de la séquence consensus nucléique A3 et A7 de ces domaines (voir tableau 1). Ils ont ainsi détecté à l'aide de ces amorces des gènes de type NRPS dans une collection de 210 actinomycètes comme ceux de la streptogramine chez *Streptomyces virginiae* ATCC 13161 et de la pristinamycine (molécules hybride NRPS/PKS) chez *Streptomyces pristinaespiralis* DSM 40338. Plus récemment, Ramarathnam *et al.* (2007) ont défini quant à eux des amorces non dégénérées spécifiques pour la détection de la fengycine D1 et de la bacillomycine D. Ils ont, en effet, à partir des séquences nucléiques des opérons responsables de la synthèse de ces molécules, déterminé des amorces non dégénérées en choisissant des oligonucléotides présents à l'aide d'un logiciel Primer3^R (Rozen and Skaletsky 2000). Ces auteurs ont ainsi, à l'aide de ces amorces mis

en évidence des gènes de lipopeptides tels que ceux de: la fengycine chez *Bacillus subtilis* DFHO8 (souche isolée du blé, *Triticum spp*) et la bacillomycine D chez *B. cereus* DFE4.

Ces techniques restent limitées dans le premier cas à la traduction des motifs protéiques en acides nucléiques pour les PCR alors que la seconde approche utilisant des domaines d'adénylation semble non spécifique aux synthétases non ribosomiales, ce qui a d'ailleurs conduit à l'amplification de gènes non NRPS lors du criblage de la collection des actinomycètes. Ces deux approches moléculaires utilisant l'amplification par PCR aboutissent à la détection de petits fragments de taille généralement inférieure à 1000 bp.

6.2. Approches prédictives bioinformatiques

Aujourd'hui, la production de très importantes données de séquences couplée au développement d'outils bioinformatiques pour l'analyse de ces séquences ont permis de progresser sur la connaissance du génome bactérien. Ces études ont, en outre, révélé que les transferts horizontaux d'information génétique jouent un rôle prépondérant dans la plasticité des génomes bactériens. De nombreux outils tant basés sur des méthodes informatiques d'analyses de séquences que sur des expérimentations au niveau moléculaire ont été mis au point.

Des expériences de mutagenèse dirigée appliquées sur des gènes NRPS ont permis de démontrer la pertinence du choix des acides aminés clefs ou signatures des domaines d'adénylation. En effet, Eppelmann *et al.* (2002) ont montré qu'une seule mutation dans le domaine adénylation induit la reconnaissance spécifique d'un autre substrat muté sans perte d'efficacité.

Des outils informatiques permettent quant à eux d'établir un agencement structural des différents domaines composant un opéron impliqué dans la synthèse non ribosomiale par une reconnaissance de motifs spécifiques.

Certains logiciels permettent également de prédire le substrat que peut activer le domaine d'adénylation des synthétases non ribosomiales. En effet, des études basées sur les données structurales de domaine d'adénylation, dont la structure tridimensionnelle a été déterminée par cristallisation et diffraction des rayons X, ont permis d'identifier les acides aminés d'activation des domaines A permettant ainsi de modéliser le substrat susceptible d'être reconnu et activé. La détermination de tous les substrats au sein d'un opéron permettrait non seulement de trouver la composition en substrats de la molécule mais aussi de prédire sa structure et de là, dans certains cas, sa fonction.

Une autre stratégie consistant à déterminer le substrat incorporé est basée sur la phylogénie des domaines A qui est la conséquence des transferts horizontaux de gènes (THG).

L'intérêt de ces outils réside dans le fait qu'ils permettent d'aider le biologiste à orienter ces expérimentations et surtout de prédire les structures des fonctions biologiques des molécules sans avoir recours à des analyses très coûteuses.

On peut regrouper ces outils en deux catégories distinctes bien qu'ils soient complémentaires.

6.2.1. Outils permettant la détermination du substrat activé

En étudiant la structure moléculaire du domaine A du premier module activant la Phe de la gramicidine synthétase, Conti *et al.* (1997) ont déterminé 10 motifs protéiques conservés (voir Tableau 1). Ces auteurs ont montré que la spécificité de l'activation des substrats est déterminée par le segment protéique A3-A6 alors que plus tard, Schwarzer *et al.* (2003) ont affirmé que cette spécificité est le fait du segment A4-A5. Sur la base de la structure cristalline du premier domaine A de la gramicidine synthétase, Stachelhaus *et al.* (1999) et Challis *et al.* (2000) ont défini le site actif de l'enzyme. Ils ont identifié dans ce dernier, 10 acides aminés qui sont directement impliqués dans la reconnaissance de l'acide aminé activé. Un alignement multiple de différents domaines d'adénylation pour lesquels la spécificité du substrat est connue les a conduits à définir un code conférant une spécificité ou code de Stachelhaus « specificity-conferring code ». Les bases de données de cet outil proviennent essentiellement des domaines A des souches *E. coli*, *Bacillus spp.* et *Pseudomonas spp.* Cette base de données élude un groupe important de producteurs de molécules non ribosomiales que sont les actinomycètes, principalement le genre *Streptomyces*.

Rausch *et al.* (2005) ont alors déterminé à partir d'un alignement de 397 différents domaines d'adénylation un ensemble de résidus situés à 8Å du site actif.

En comparant les prédictions pour 1230 domaines A de molécules NRPS de la banque protéique UniProt, la nouvelle méthode (celle de Rausch) a montré une fiabilité supérieure de 18% comparée à celle du code de Stachelhaus dans la détermination de la spécificité des substrats activés.

Ces bases de données s'enrichissent régulièrement mais l'expérience prouve que la complexité des signatures ne permet toujours pas de déterminer la nature du substrat incorporé comme dans le cas des gènes NRPS de *Lactobacillus plantarum* WCSF1 (voir Chapitre 5).

Tous ces outils restent limités pour prédire de façon objective la nature de la molécule finale essentiellement car ils sont basés sur des opérons appartenant à un nombre restreint d'espèces.

6.2.2. Outils permettant la détermination de l'agencement structural des opérons

La détermination des core motifs protéiques a été étendue sur les autres domaines NRPS (C, A, T, E, N, Ox) qui sont alors utilisés pour leur identification (Schwarzer *et al.* 2003).

Ansari *et al.* (2004), en se basant sur l'ensemble de ces signatures, ont mis en place un logiciel appelé NRPS/PKS, qui permet de prévoir l'organisation en modules et en domaines des synthétases non ribosomiales et polykétides synthases. Ce logiciel fournit une interface conviviale pour identifier les structures des enzymes modulaires. Ils se sont basés sur la comparaison de la séquence cible (séquence à analyser) et 307 ORFs, 2223 différents domaines dont 101 structures de molécules différentes.

De même, les travaux sur les signatures des différents domaines ont permis de créer plusieurs outils non commerciaux comme le logiciel NRPS de l'équipe du Pr. Philippe Jacques, ProBiogem, France (non publié) qui permet d'identifier les différents domaines NRPS d'une protéine, mais aussi des outils commerciaux (Vector NTI, ClustScan). Le développement de ces outils peut être corrélé à l'importance de ces molécules dans différents secteurs d'activités.

L'avantage est sans nul doute l'amélioration de l'efficacité des découpages ou la détermination des limites de chaque domaine à l'intérieur de chaque synthétase.

Chapitre 2 Les lipopeptides: des peptides non ribosomiaux de *Bacillus*

Les lipopeptides sont des molécules composées d'un acide gras (β -hydroxy ou β -amino) et d'une chaîne peptidique qui peut être linéaire, branchée ou cyclique. La plupart des molécules lipopeptidiques produits par le genre *Bacillus* sont subdivisées en famille, en fonction de leur structure. On peut néanmoins classer ces lipopeptides en 2 catégories :

- Catégorie I : les lipopeptides produits par le genre *Bacillus* selon le mécanisme non ribosomal.

Cette catégorie de lipopeptides englobe la plupart des lipopeptides produits par le genre *Bacillus* (*B. subtilis*, *coagulans*, *amyloliquefaciens*, *pumilus*, *cereus*, *thuringiensis*, *brevis* et *licheniformis*). Ils sont divisés en 3 familles (les fengycines, les surfactines et les iturines) selon le niveau d'homologie entre les synthétases et des acides aminés constitutifs des molécules (Zuber *et al.* 1993, Ongena and Jacques 2008). Ces lipopeptides sont d'ailleurs les plus étudiés.

- Catégorie II. Les lipopeptides dont le mécanisme non ribosomal de synthèse n'était pas encore démontré en début de ce travail.

Il s'agit de certains lipopeptides comme les kurstakines produits par *Bacillus thuringiensis* (Hathout *et al.* 2000, Price *et al.* 2007).

D'autres lipopeptides non ribosomiaux sont produits par d'autres genres comme les daptomycines (antibiotique lipopeptidique) produites par le genre *Streptomyces* (Doekel *et al.* 2008; Hachmann *et al.* 2009).

Notre intérêt s'est porté dans ce chapitre sur les lipopeptides de la catégorie I c'est-à-dire les lipopeptides-NRPS issus du genre *Bacillus*. Ces molécules font l'objet de nombreuses études au sein du laboratoire ProBioGEM. Elles sont synthétisées de façon linéaire et possèdent plusieurs activités biologiques.

I. Structures et activités biologiques des lipopeptides

Ces molécules renferment une partie peptidique composée de 7 ou 10 résidus d'acides α -aminés L ou D, reliés à un acide gras β -hydroxylé ou β -aminé à longue chaîne carbonée. Pour chaque famille, on distingue des variants se différenciant par la nature des acides aminés dans le cycle, et des homologues qui varient suivant la longueur (de 13 à 18 atomes de carbone) et l'isomérisation (linéaire, iso ou antéiso) de la chaîne lipidique. Plusieurs auteurs ont montré que les activités biologiques des lipopeptides sont étroitement liées à leurs structures. Une caractéristique principale des lipopeptides des souches de *Bacillus subtilis*, réside dans leurs propriétés

tensioactives en raison de leur structure amphiphile caractérisée par une partie polaire et une chaîne hydrocarbonée hydrophobe. Ces lipopeptides sont en effet susceptibles de diminuer la tension de surface, de modifier et de perturber les bicouches lipidiques (Deleu *et al.* 2003; Heerklotz and Seelig 2007).

Le tableau 2 montre la structure primaire des lipopeptides produits par le genre *Bacillus*.

Tableau 2. Structure primaire de quelques lipopeptides de *Bacillus*

Lipopeptides	Structures	Souches productrices	Auteurs
Lichenysines	FA α -L-Glu(Gln)-L-Leu(Ile)-D-Leu-L-Val(L-Ile)-L-Asp-D-Leu-L-Ile(L-Val)	<i>B. licheniformis</i>	Ongena and Jacques (2008)
Surfactines	FA α -L-Glu-L-Leu(Val,Ile)-D-Leu-L-Ala(Val,Leu,Ile)-L-Asp-D-Leu-L-Leu(Val,Ile)	<i>B. subtilis</i>	Ongena and Jacques (2008)
Espérine	FAB β -L-Glu-L-Leu-D-Leu-L-Val-L-Asp-D-Leu-L-Leu(L-Val)	<i>B. subtilis</i>	Thomas and Ito (1969)
Pumilacidine	FA α -L-Glu-L-Leu-D-Leu-L-Leu-L-Asp-D-Leu-L-Val(L-Ile)	<i>B. pumilis</i>	Ongena and Jacques (2008)
Iturine A	FAc-L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Asn-L-Ser	<i>B. subtilis</i>	Peypoux <i>et al.</i> (1978)
Iturine C	FAc-L-Asp-D-Tyr-D-Asn-L-Gln-L-Pro-D-Asn-L-Ser	<i>B. subtilis</i>	Ongena and Jacques (2008)
Bacillomycine L	FAc-L-Asp-D-Tyr-D-Asn-L-Ser-L-Gln-D-Ser-L-Thr	<i>B. subtilis</i>	Ongena and Jacques (2008)
Bacillomycine D	FAc-L-Asn-D-Tyr-D-Asn-L-Pro-L-Glu-D-Ser-L-Thr	<i>B. subtilis</i> , <i>B. amyloliquefaciens</i>	Ongena and Jacques (2008)
Bacillomycine F	FAc-L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Asn-L-Thr	<i>B. subtilis</i>	Ongena and Jacques (2008)
Mycosubtiline	FAc-L-Asn-D-Tyr-D-Asn-L-Gln-L-Pro-D-Ser-L-Asn	<i>B. subtilis</i>	Ongena and Jacques (2008)
*Fengycines A (B)	FA δ -L-Glu-D-Orn-D-Tyr-D-alloThr-L-Glu-D-Ala(D-Val)-L-Pro-L-Gln-L-Tyr-L-Ile	<i>B. subtilis</i>	Ongena and Jacques (2008)
*Plipastatines A (B)	FA δ -L-Glu-D-Orn-L-Tyr-D-alloThr-L-Glu-D-Val(D-Ala)-L-Pro-L-Gln-D-Tyr-L-Ile	<i>B. subtilis</i>	Ongena and Jacques (2008)

* double liaison entre les carbones 2-3, 3-4 et 13-14 de la chaîne acide gras

FA α , β -hydroxy acide gras. Le groupe β -hydroxy forme un pont ester avec le groupe carboxyl à l'extrémité C-terminale de l'acide aminé.

FAB β , β -hydroxy acide gras. Le groupe β -hydroxy forme un pont ester avec le groupe carboxyl de Asp en position 5.

FAc, β -amino acide gras. Le groupe β -amino forme un pont peptidique avec le groupe carboxyl de l'extrémité C-terminale de l'acide aminé.

FA δ , β -hydroxy acide gras. Le groupe carboxyl à l'extrémité C-terminale forme un pont ester avec le groupe hydroxyl de la Tyr en position 3.

La surfactine, la lichenysine, la pumilacidine et l'espérine appartiennent à la même famille lipopeptidique selon Ongena and Jacques (2008). Ce sont des heptapeptides reliés à des acides gras β -hydroxylés formant un pont ester soit avec le carbone C-terminal de la chaîne peptidique pour la surfactine, la lichenysine et la pumilacidine ou soit avec le groupe carboxyl de l'Asp en position 5 pour l'espérine. La surfactine et la lichenysine possèdent un pouvoir biosurfactant et des propriétés hémolytiques limitant fortement leur exploitation potentielle en thérapeutique humaine. En outre, la surfactine produite par *B. subtilis* est un bioémulsifiant biologique très efficace. Elle stabilise les mousses formées en sa présence à cause de ces caractéristiques d'hydrophobie due à sa chaîne alkyl et ces propriétés moléculaires intrinsèques. Razafindralambo *et al.* (1998) ont montré que la meilleure capacité à former des mousses stables

a été obtenue avec des chaînes de surfactine possédant quatorze atomes de carbone. La surfactine est également capable d'interagir avec des biomembranes (Deleu *et al.* 1999; Grau *et al.* 1999). Cette activité biosurfactante est meilleure à pH alcalin qu'à pH acide (Morikawa *et al.* 2000). Cette molécule a, en outre, des activités antivirale et antibactérienne en particulier antimycoplasme (Kracht *et al.* 1999).

La lichenysine, quant à elle, est produite par *Bacillus licheniformis* et décrite comme possédant un pouvoir surfactant supérieur à celui de la surfactine (Grangemard *et al.* 2001). La lichenysine est par ailleurs un très bon agent chélatant des cations.

Des composés tels que la pumilacidine, une molécule antivirale (Naruse *et al.* 1990) et l'espérine (Konz *et al.* 1999) produites respectivement par *B. pumilis* et *B. subtilis* sont considérées comme des variants appartenant à la famille des surfactines (Ongena and Jacques 2008).

Les molécules de la famille des fengycines sont des lipodecapeptides comportant une chaîne β -hydroxy d'acide gras (Steller *et al.* 1999). Cette famille regroupe essentiellement deux types de molécules que sont les plipastatines et les fengycines.

La distinction entre ces deux lipopeptides réside dans la stéréochimie du résidu Tyr en position 3 de la chaîne peptidique, qui est de conformation L pour les plipastatines et D pour les fengycines (Umezawa *et al.* 1986; Nishikiori *et al.* 2006). Les séquences des promoteurs *fenC* et de *ppsA* diffèrent aussi d'un nucléotide au niveau du 10^{ème} domaine: TATAAT dans le promoteur *fenC* et TATAGT pour le promoteur *pps*. Lin *et al.* (1999) ont montré que ce changement unique a un rôle prépondérant dans la transcription du gène *pps*. Toutefois Steller *et al.* (1999), en caractérisant les gènes *pps* et *fen*, ont montré leur forte homologie supérieure à 70%.

Il semblerait que ces molécules aient en général les mêmes fonctions. Pourtant, Umezawa *et al.* (1986) ont montré l'action inhibitrice des plipastatines sur la phospholipase A₂ (PLA₂) de *Bacillus cereus* BMG302-fF67 contrairement aux fengycines qui auraient un effet stimulateur. L'inhibition de la PLA₂ par les plipastatines a été confirmée par Volpon *et al.* (2000). On ignore cependant si cela est lié à la différence des deux molécules au niveau de la stéréochimie du résidu Tyr.

En général, les fengycines sont essentiellement des composés antifongiques notamment contre *Fusarium moniliforme* Sheldon ATCC 38932 (Hu *et al.* 2007). Des travaux ont déjà démontré les effets protecteurs de la fengycine produite par la souche *B. subtilis* S499, dans certains pathosystèmes tels que le haricot/*Pythium ultimum*, le concombre/*Colletotrichum lagenarium* et la tomate/*Pythium aphanidermatum* (Ongena *et al.* 2005a; Ongena *et al.* 2005b).

La famille des iturines comprend le plus grand nombre de molécules et de variants parmi les lipopeptides ayant été étudiés à ce jour. Elle comprend les iturines A et C, la bacillomycine D, F,

2.1. Opéron des surfactines

L'opéron surfactine qui regroupe l'ensemble des gènes codant les synthétases impliquées dans la synthèse de la surfactine a été mis en évidence chez *Bacillus subtilis* 168 (Kunst *et al.* 1997). Il est constitué de 4 gènes (*srfA-A*, *srfA-B*, *srfA-C* et *srfA-D*) qui codent respectivement les protéines SrfA-A, SrfA-B, SrfA-C et SrfA-D comme présenté à la figure 14. Excepté la synthétase SrfA-D, les autres jouent un rôle actif lors de cette biosynthèse. En effet, la synthétase SrfA-D est constituée uniquement d'un domaine Te (intervenant dans le clivage de la molécule néoformée), on parle de module tronqué. La figure 15 présente les principaux modules et les domaines constitutifs des synthétases fonctionnels de l'opéron de la surfactine.

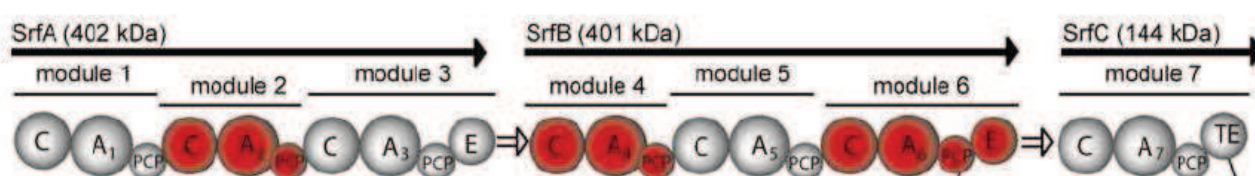


Figure 15. Opéron de la surfactine synthétase (Sieber and Marahiel 2005)

Ce complexe multienzymatique est composé de 7 modules spécifiques pour l'incorporation de 7 acides aminés.

La synthétase SrfA-A (3588 aa, 402 KDA) incorpore successivement les acides aminés L-Glu, L-Leu et D-Leu, SrfA-B (3587 aa, 401 KDA) L-Val, L-Asp et D-Leu et SrfA-C (1274 aa, 144 KDA) L-Leu.

Les protéines SrfA-A, SrfA-B et SrfA-C de la surfactine synthétase sont codées respectivement par les gènes *srfA-A* (10767 bp), *srfA-B* (10764 bp) et *srfA-C* (3825 bp).

La lichenysine synthétase qui synthétise la lichenysine, molécule produite par *Bacillus licheniformis* et appartenant à la famille des surfactines (Ongena and Jacques 2008), a une organisation opéronique similaire à celle de la surfactine. En effet, la lichenysine synthétase est aussi composée des ORFs *lchA*, *lchB*, *lchC* et *lchD* (voir figure 14).

2.2. Opéron des fengycines

Les fengycines comprennent deux types de molécules : les plipastatines et les fengycines.

L'opéron de la plipastatine de *Bacillus subtilis* 168 a été complètement séquencé par Kunst *et al.* (1997). Son organisation est présentée dans la figure 14. La synthétase de la plipastatine est codée par le gène *pps* de taille 37689 bp et, est répertorié dans la banque Genbank sous le numéro d'accèsion Z99113. Le gène *pps* est flanqué en début des gènes *dacC* et *yoxA* puis à la fin du gène *yngL*. La synthétase de la plipastatine est constituée de 5 protéines : PpsA (2561 aa,

289 kDA), PpsB (2560 aa, 290 kDA), PpsC (2555 aa, 287 kDA), PpsD (3603 aa, 407 kDA) et PpsE (1279 aa, 145 kDA) codées respectivement par les gènes *ppsA* (7686 bp), *ppsB* (7683 bp), *ppsC* (7668 bp), *ppsD* (10812 bp) et *ppsE* (3840 bp). Les synthétases PpsA, PpsB et PpsC sont chacune constituées de 2 modules alors que PpsD et PpsE contiennent respectivement 3 et 1 modules. Chaque synthétase est responsable de l'incorporation des acides aminés suivants: L-Glu et D-Orn pour PpsA, L-Tyr et D-allo-Thr pour PpsB, L-Glu et D-Val/Ala pour PpsC, L-Pro, L-Gln et D-Tyr pour PpsD et L-Ile pour PpsE.

La fengycine appartient à la même famille que la plipastatine, l'organisation de sa synthétase est similaire à celle de cette dernière et, est constituée des protéines FenC, FenD, FenE, FenA et FenB (Lin *et al.* 1999).

2.3. Opéron des iturines

Contrairement aux autres synthétases impliquées dans la synthèse des lipopeptides, les synthétases d'iturines sont des produits issus d'une combinaison de deux systèmes de synthèses de métabolites secondaires différents: les Polykétides Synthases (PKS) et les NRPS (voir figure 14).

Plusieurs auteurs (Duitman *et al.* 1999, Leclère *et al.* 2005) ont décrit la synthétase de mycosubtiline comme étant constituée de 4 protéines : FenF, MycA, MycB et MycC qui sont respectivement codées par les gènes *fenF* (1200 bp), *mycA* (11916 bp), *mycB* (16110 bp) et *mycC* (7830 bp). La synthétase MycA (3971 aa, 449.3 kDA) permet l'incorporation dans la molécule finale d'une L-Asn, la MycB (5369 aa, 612.3 kDA) une D-Tyr, un D-Asn, un L-Gln et un L-Pro et MycC (2609 aa, 297.9 kDA) une D-Ser et une L-Asn.

La structuration des synthétases de l'iturineA (Tsuge *et al.* 2001) et de la bacillomycine D (Moyne *et al.* 2004) est similaire à celle de la mycosubtiline.

En général, les synthétases impliquées dans la synthèse de molécules d'une même famille ont une organisation structurale semblable et une très faible divergence au niveau des substrats activés. Les différences au niveau des acides aminés constitutifs des molécules de lipopeptides sont essentiellement le fait des changements intragéniques intervenus entre les domaines au niveau des synthétases, ce qui induit par exemple l'inversion des deux derniers acides aminés au niveau de MycC et ItuC.

- **Conclusions**

De par leur nature biologique, les lipopeptides constituent une alternative aux molécules chimiques pour l'environnement, l'agriculture et la médecine.

De nos jours, la production de lipopeptides est confinée à une petite échelle bien que de nombreux essais pour leur production sélective à l'échelle industrielle soient réalisés. Elle reste néanmoins limitée du fait de la difficulté de la gestion de la mousse que les lipopeptides génèrent au cours de leur production et de leur purification.

Chapitre 3 But du travail

Les applications potentielles des molécules d'origine non ribosomiale sont réelles et suscitent un intérêt croissant pour la recherche (mise en évidence de nouvelles structures peptidiques) et aussi pour certaines industries notamment agro-alimentaire, pharmaceutique, chimique etc...

L'objectif de cette étude est d'une part, de mettre au point des méthodes de détection de souches productrices de ces molécules par analyse moléculaire des gènes impliqués dans leur synthèse et d'autre part, d'identifier les molécules associées.

La finalité de ce travail étant d'aboutir à terme, à la mise au point d'outils capables de renseigner sur le potentiel NRPS des différentes souches de microorganismes.

Il s'avère alors intéressant de rechercher les molécules NRPS chez ces souches potentiellement productrices par une stratégie permettant de pallier les difficultés de leur recherche par des analyses biochimiques. En effet ces molécules non ribosomiales sont recherchées à partir des surnageants de culture ou de colonies bactériennes.

Néanmoins, cette technique n'a pas toujours permis de détecter les capacités de synthèse des bactéries soit parce que les procédés de culture mise en œuvre ne sont pas toujours optimisés soit parce que le seuil de détection des méthodes analytiques peut s'avérer être une limite. Cela a été le cas de la souche *Bacillus subtilis* S499, connue jusqu'alors uniquement pour sa production de la surfactine S1 et de l'iturine A. L'optimisation des procédés de fermentation de cette souche S499 a non seulement permis l'augmentation de la production de cette surfactine mais a aussi permis de révéler la présence d'une troisième famille de lipopeptides: les fengycines produites pendant la phase stationnaire de croissance (Jacques *et al.* 1999).

Une autre limite de cette méthode est liée au phénotype bactérien. Lorsque la molécule recherchée n'est pas exprimée ou l'est en faibles quantités par la souche, les analyses biochimiques s'avèrent inefficaces. C'est le cas de *B. subtilis* 168 qui bien que possédant les opérons surfactine et plipastatine est incapable de les synthétiser à cause d'une protéine Sfp non fonctionnelle.

L'alternative aux méthodes analytiques est le criblage moléculaire qui consiste à la mise en évidence des gènes impliqués dans la synthèse de molécules non ribosomiales.

En effet, ces techniques permettent de mettre en évidence des molécules NRPS par la présence de leurs gènes. Elles permettent ainsi de contourner les difficultés liées aux méthodes analytiques comme décrites précédemment.

Le cadre de ce travail consiste à rechercher des molécules non ribosomiales par la détection des gènes impliqués dans leur synthèse par PCR. Il s'agira ainsi de définir une approche innovante

par la détermination des amorces capables à la fois d'être une alternative fiable aux difficultés liées aux méthodes analytiques, mais aussi à ceux liés à l'utilisation des amorces définies à partir des séquences consensus protéiques.

Nous présentons sous formes de chapitres, l'ensemble de nos travaux réalisés pour la recherche de molécules non ribosomiales.

Les trois premiers chapitres des résultats sont consacrés à la mise au point de la technique de détection de molécules NRPS à partir des gènes de microorganismes modèles, de l'application de cette technique à différentes collections de bactéries lactiques et de bacilles issus de produits fermentés africains. Le quatrième chapitre sera consacré à la conception d'un logiciel de bioinformatique (PrimerDeg 0.1) qui permet l'automatisation des procédures pour définir des amorces spécifiques. Le dernier chapitre traitera de la combinaison de cette stratégie moléculaire aux outils classiques biochimiques et bioinformatiques pour l'identification d'une nouvelle molécule non ribosomiale.

PARTIE II.
EXPERIMENTATIONS

Chapitre 4 Matériels et Méthodes

I. Matériels

1.1. Milieux de culture

Les milieux de culture ont été préparés dans des conditions définies selon les recommandations des auteurs ou des différents fournisseurs.

1.1.1. Milieu de Man–Rogosa–Sharpe ou MRS

Le milieu MRS est préparé en diluant 55,3 g de bouillon MRS (Biokar Diagnostics, Beauvais, France) dans 1 L d'eau stérile. La solution ainsi obtenue est autoclavée à 121°C pendant 30 min. Le milieu solide MRS est obtenu par l'ajout de 1,7% d'agar dans le bouillon ainsi préparé avant la stérilisation.

1.1.2. Milieu de Luria-Bertani ou LB

Le milieu Luria-Bertani (LB) est préparé avec 10 g de tryptone, 5 g d'extrait de levure et 10 g de NaCl. Son pH est ajusté à 7.0-7.2 à l'aide de NaOH puis stérilisé à l'autoclave pendant 20 min à 121°C. Un ajout de 1,7% d'agar est effectué pour l'obtention du milieu solide correspondant.

1.1.3. Milieu DSM

La réalisation du milieu DSM nécessite le mélange de deux solutions S1 solide et S2 liquide.

S1 est préparé dans un ½ L d'eau MilliQ et contient 8 g de bouillon nutritif, 0,108 g de MgSO_4 heptahydraté, 1,8g de KCl et 8,5g d'agar.

La solution S2 contient 236,15 mg de $\text{Ca}(\text{NO}_3)_2$ tétrahydraté, 0,198 mg de MnCl_2 tétrahydraté et 0,0278 mg de FeSO_4 heptahydraté dans un ½ L d'eau MilliQ.

Les deux milieux sont autoclavés à 121°C pendant 30 min et mélangés stérilement avant utilisation.

Le milieu DSM a été mis au point pour la production sélective de certains lipopeptides.

1.1.4. Milieu Potato Dextrose Agar ou PDA

Le milieu PDA est constitué de 4 g de pomme de terre, 10 g de glucose et 15 g d'agar. Sa préparation est réalisée par la dissolution de 20 g de bouillon déshydraté (Biokar Diagnostics, Beauvais, France) dans 1 L de solution. Cette solution ainsi obtenue est stérilisée pendant 30 min à 121°C avant utilisation.

1.1.5. Milieux 853 et 856

Le milieu 853 est composé de 10 g de tryptone, 5 g d'extrait de levure, 5 g de NaCl, 1 g de glucose et 0.7 g de K₂HPO₄ pour 1 L. Il est stérilisé à 121°C pendant 30 min avant utilisation.

Le milieu 856 est préparé par l'ajout de l'agar à 1,7% dans le milieu 853.

1.1.6. Milieux 863 et 868

Le milieu 863 est constitué de 10 g de peptone, 10 g de glucose et 10 g d'extrait de levure dissous dans 1 L d'eau. La solution ainsi obtenue est stérilisée à 121°C pendant 30 min.

Le milieu 868 est préparé à partir du milieu 863 avec l'ajout de 1,7% d'agar.

1.1.7. Milieu SOC

Le milieu SOC est composé de 20 g de tryptone, 5 g d'extrait de levure, 0,5 de NaCl et 10 mL de KCl 250 mM (1,86 g/100 mL) pour 1 L de solution.

Son pH est ajusté à 7 avec une solution de NaOH puis stérilisé à l'autoclave à 110°C pendant 30 min.

1.1.8. Milieu nutritif

Le milieu nutritif ou Bouillon Nutritif (BN) est préparé à partir de 20 g de bouillon nutritif déshydraté (Biokar Diagnostics Beauvais, France) dissous dans 1 L d'H₂O. Il est composé de 10 g de tryptone, 5 g d'extraits de viande et 5 g de chlorure de sodium.

Le milieu solide équivalent appelé Gélose Nutritive (GN) est complémenté avec 1,7% d'agar.

1.1.9. Eau peptonée

1 L d'eau peptonée est composé de 40 g de peptone, 20 g de NaCl et 1 mL de Tween 80.

1.2. Souches bactériennes

1.2.1. Souches de *Lactobacillus*

Lactobacillus plantarum WCSF1 est une souche dérivée de *L. plantarum* NCIMB8826 (Vesa *et al.* 2000). Son génome est complètement séquencé et annoté (Kleerebezem *et al.* 2003), avec un numéro d'accèsion Genbank AL935263. L'analyse de ce génome montre la présence des gènes *npsA* et *npsB* codant pour des synthétases impliquées dans la synthèse non ribosomiale. Cette souche a servi de référence pour l'étude des souches du genre *Lactobacillus*.

D'autres *Lactobacillus* d'origines diverses ont été utilisés. Ils sont présentés dans le tableau 3. Jusqu'à présent, aucune autre souche de lactobacilles n'est connue comme possédant des gènes NRPS.

Tableau 3. Caractéristiques des souches de *Lactobacillus* étudiées

Souches	Origines
<i>Lactobacillus plantarum</i> WCSF1	NIZO, Pays-Bas
<i>Lactobacillus plantarum</i> NCIMB 8826	Offert par IBL, France
<i>Lactobacillus plantarum</i> NZ 1400	NIZO, Pays-Bas
<i>Lactobacillus sakei-sakei</i> 16	ProBioGEM
<i>Lactobacillus plantarum</i> ATCC 14917	BCCM/LMG, Belgique
<i>Lactobacillus plantarum</i> ATCC 8014	BCCM/LMG, Belgique

Lactobacillus sakei-sakei 16 (Lb16) de la collection de notre laboratoire a été isolé de hareng fumé. Cette souche possède une activité antifongique (anti-*Saccharomyces cerevisiae*, anti-*Pichia pastoris*, anti-*Penicillium roquefortii* et anti-*Mucor rouxii*) comparable à celle de *L. plantarum* 8826 (dont est issu le clone WCSF1). La souche *L. plantarum* NZ 1400 est un clone de la souche WCSF1, son gène *npsA* a été inactivé.

Les souches de *L. plantarum* ATCC 14917, ATCC 8014 de la collection BCCM/LMG ont été décrites comme ne possédant pas de gènes NRPS (Moleenar *et al.* 2005).

D'autres souches de bactéries lactiques ont été isolées dans cette étude. Elles proviennent de *Clupea harengus* (hareng) fumé. Ces souches forment une collection composée de plusieurs espèces telles *L. plantarum*, *L. casei*, *L. sakei-sakei*, *L. pentosis*, *L. sakei-curvatus*, *L. pentosus* et aussi d'autre genre tel *Lactococcus lactis*.

Les lactobacilles étant anaérobies facultatifs, leur croissance a été réalisée dans le milieu MRS à 30°C pendant 24 h, dans une jarre enrichie en CO₂ par la réaction entre l'acide citrique et le bicarbonate de sodium.

1.2.2. Souches de *Bacillus*

Les souches de *Bacillus* ont été choisies pour leur potentiel de production ou non de différents lipopeptides. Elles ont été incubées dans le milieu Luria-Bertani (LB), en aérobie pendant 24 h à 30°C.

- *Bacillus subtilis* 168

La souche de *Bacillus subtilis* 168 utilisée est issue de la collection « Bacillus Stock Genetic Center », son génome est complètement séquencé (Kunst *et al.* 1997).

L'annotation du génome indique la présence des opérons *surfA* et *pps* codant respectivement pour les synthétases de surfactine et de plipastatine. Cette souche a été utilisée comme souche de

référence pour la présence des gènes impliqués dans la synthèse des lipopeptides : surfactine et plipastatine.

- *Bacillus subtilis* ATCC 21332, ATCC 9943, ATCC 6633

Ces souches ont été utilisées car elles sont décrites comme productrices de peptides spécifiques: La surfactine par les 3 souches (Kluge *et al.* 1988; Duitman *et al.* 1999), l'iturine et la fengycine par ATCC 9943 (Vater *et al.* 2002; Chollet-Imbert *et al.* 2009), la mycosubtiline par ATCC 6633 (Duitman *et al.* 1999; Leclère *et al.* 2005).

- *Bacillus licheniformis* ATCC 14580

Le génome de *B. licheniformis* ATCC 14580 a été complètement séquencé et annoté (Rey *et al.* 2004). On le retrouve dans Genbank sous le numéro d'accès CP000002. Il contient plusieurs gènes impliqués dans la synthèse d'un lipopeptide : la lichenysine. D'autres espèces comme ATCC 10716 sont décrites pour la synthèse de d'autres molécules NRPS comme la bacitracine (Haavik 1974). Uniquement l'espèce ATCC 14580 de la collection de notre laboratoire a été utilisée lors de cette étude, elle nous a servi comme contrôle négatif pour la synthèse de certains lipopeptides.

- *Bacillus thuringiensis*

Une souche, *Bacillus thuringiensis* serovar *berliner* 1915 (Institut Pasteur de Paris, CRBIP France) et une autre, *Bacillus thuringiensis* S 45 isolée de céréales, offerte par l'université d'Alger (Algérie), ont été utilisées pour rechercher d'éventuels potentiels génétiques NRPS car elles possèdent des activités biologiques.

- *Bacillus cereus* LMG 2098

La souche de *B. cereus* LMG 2098 nous a été fournie par le « Laboratorium voor Microbiologie » de Ghent en Belgique. Elle nous a intéressé car elle n'est pas connue comme étant capable de produire des NRPS de type lipopeptide.

A côté de ces souches de bacilles décrites ci-dessus, d'autres souches d'origines diverses ont été criblées pour la recherche de molécules NRPS.

En effet, différentes souches de *Bacillus* ont été isolées dans cette étude à partir de produits fermentés à base de céréales provenant du Burkina-Faso (Afrique de l'Ouest). Quelques-unes ont été retenues sur la base de la présence d'activités biologiques.

1.2.3. Autres souches

- *Micrococcus luteus*

La souche de *Micrococcus luteus* est une bactérie Gram+ à coques, mesurant 0,5 à 2 µm de diamètre, appartenant à la famille des *Micrococaceae* généralement immobiles excepté l'espèce *agilis*. Elle a été utilisée pour des tests antimicrobiens.

- Moisissures et levures

Certains champignons ont été utilisés pour des tests antifongiques. Il s'agit des souches de *Rhizopus nigricans*, *Penicillium roquefortii*, *Mucor rouxii*, *Fusarium oxysporum* et *Aspergillus niger* comme moisissures, *Saccharomyces cerevisiae* comme levure. La croissance des souches de moisissures a été réalisée sur milieu liquide et solide PDA (Potato Dextrose Agar) contenant du chloramphénicol (concentré à 0,5 g/L) à 20°C pendant une dizaine de jours. Les souches de levures ont été incubées sur un milieu 868 pendant une nuit.

- *Escherichia coli*

Deux souches d'*E. coli*, l'espèce *E. coli* JM109 de Promega Corp. (USA) et XL1-Blue de Fermentas (France) ont été utilisées comme cellules compétentes pour l'amplification et la conservation des vecteurs plasmidiques.

Les génotypes sont:

JM109 : *recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17*, *supE44*, *relA1*, $\Delta(lac-proAB)/F'$ [*traD36*, *proAB*⁺, *lacI*^q, *lacZ* Δ M15],

XL1-Blue : *recA*, *endA1*, *gyrA96*, *thi*, *hsdR17* (*rk*-,*mk*+), *supE44*, *relA1*, *lambda*-, *lac*-, [*F'*, *proAB*, *lacIqZdeltaM15*, *Tn10(tet)*].

Les cellules compétentes d'*E. coli* JM109 ont été acquises commercialement tandis que celles de XL1-Blue ont été fournies gracieusement par l'Institut de Recherches Microbiologiques Wiame de Bruxelles (Belgique). Elles ont été incubées dans du milieu LB (complémenté ou non en antibiotiques ou en composés spécifiques) pendant 24 h à 37°C dans un bain-marie sous agitation (environ 20 g).

1.3. Vecteurs plasmidiques

Deux vecteurs commerciaux ont été essentiellement utilisés lors des transformations: pJET1/blunt a été utilisé pour le clonage avec *E. coli* XL1-Blue et pGEMT-easy pour *E. coli* JM109 bien qu'ils soient tous deux adaptés à l'une ou l'autre des cellules compétentes.

Le vecteur pJET1/blunt (Fermentas) de taille égale à 3128 bp est utilisé pour le clonage des produits PCR portant des bouts francs.

La figure 16 présente la carte génétique du vecteur plasmidique pJET1/Blunt et ses différents sites de restrictions.

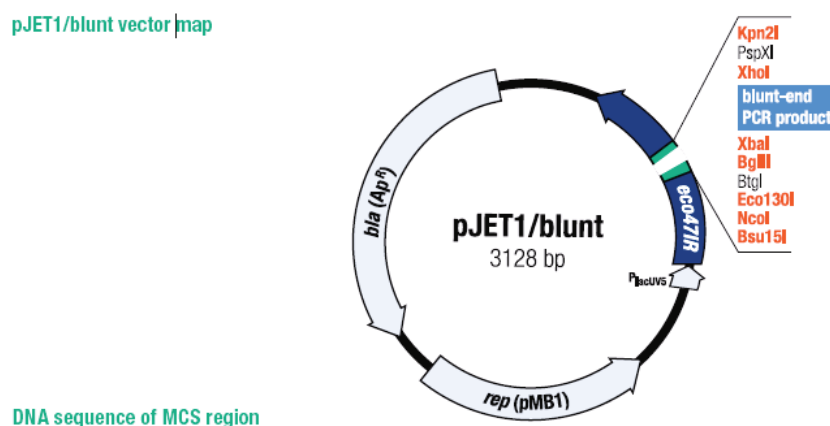


Figure 16. Carte génétique de pJET1/Blunt (Source manuel Fermentas)

Il contient un gène létal endonucléasique *eco47IR* responsable de la sélection des recombinants positifs. La modification de l'expression du gène *eco47IR* par le promoteur P_{lacUV5} est suffisante pour la sélection des transformants positifs. Il possède de plus, un gène de répllication *rep*(pMB1) et le gène *bla* (Amp^r) qui confère la résistance à l'ampicilline.

Le vecteur pGEMT-Easy (Promega Corp., USA), de taille égale à 3015 Kb a été utilisé lors de clonage des produits PCR ayant des extrémités non cohésives. Ce vecteur possède un gène de résistance à l'ampicilline (Amp^r), une origine de répllication (*ori*), une partie du gène *lacZ* codant pour l' α -peptide de la β -galactosidase et un site de clonage multiple inséré dans la partie N-terminale du gène *lacZ*.

La carte génétique du vecteur pGEMT-Easy est présentée à la figure 17.

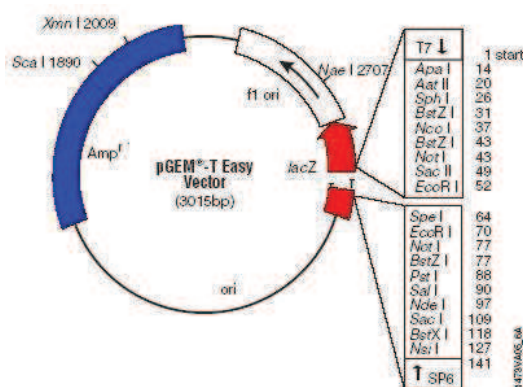


Figure 17. Carte génétique de pGEMT-Easy (Source manuel Promega)

Contrairement au vecteur pJET1/Blunt, la sélection des transformants contenant un plasmide de pGEMT-Easy nécessite de l'IPTG et de l'X-Gal dans le milieu de culture en plus de l'ampicilline.

Le choix d'un vecteur au détriment d'un autre va alors dépendre de la nature cohésive ou non des produits issus de la PCR. Cette nature dépend essentiellement du type d'enzyme polymérase utilisée lors de la PCR.

Lors de cette étude, nous avons utilisé la *MasterMix* (Fermentas) comme *Taq*-polymérase pour la PCR, qui est un mélange de produits nécessaire à la réalisation d'une PCR. Cette enzyme produit des fragments PCR avec des bouts cohésifs 3'-A ou 3'-T.

1.4. Outils bioinformatiques

Les différentes fonctions des principaux outils utilisés lors de cette étude sont détaillées dans ce paragraphe.

1.4.1. Logiciels d'analyses

- Logiciel Vector NTI Advance 10TM

Le logiciel Vector NTI Advance 10TM a été mis au point par Invitrogen (USA) pour l'analyse de séquences. C'est un outil qui nous a servi pour l'étude, l'analyse et le traitement des données de séquences obtenues expérimentalement. Il se présente sous forme d'un portail ou Workflow et offre plusieurs applications (alignement et assemblage de séquences, extraction de séquences de gènes et de protéines des banques de données, comparaisons de séquences etc..). Son utilisation sera détaillée dans les chapitres ultérieurs.

- Logiciel AmplifX 1.4.4.

AmplifX a été mis au point par Jullien *et al.* (2007) du Centre de Recherche en Neurobiologie et Neurophysiologie de Marseille (France). Il est téléchargeable gratuitement via internet. Il a permis de simuler des tests PCR sur ordinateur (test *in silico*), et a été utilisé comme banque pour conserver les séquences des amorces définies. Les conditions générales d'utilisation sont celles dites « standards » ou par « défaut ». Des conditions spécifiques sont mises au point à partir de l'onglet 'preference' du menu Edition en modifiant dans la rubrique Match, les conditions par défaut du logiciel.

- **Logiciel NRPS (Laboratoire ProBioGEM)**

Le logiciel NRPS est un outil dédié à la reconnaissance des différents domaines de synthétases et des sites actifs basés sur la signature des motifs protéiques conservés. Il a permis le découpage des domaines obtenus par d'autres logiciels. Le logiciel NRPS-ProBioGEM n'est pas encore accessible à tous car des améliorations sont en cours.

1.4.2. Sites Webs

- **National Center for Biotechnology Information**

Le National Center for Biotechnology Information (NCBI) est un institut américain dédié à l'information en biologie moléculaire. Il est l'un des sites les plus consultés par les biologistes moléculaires car disposant de plusieurs et importantes banques de données. Selon ces auteurs, deux nouvelles séquences sur trois sont d'abord soumises dans ces banques que sont Genbank, RefSeq, dBSNP. Il développe des données de biologie calculatoire, met au point des logiciels d'analyse et fournit des informations biomédicales.

C'est un portail gratuit qui offre plusieurs entrées et la plupart des logiciels d'analyses de séquences sont reliés à ses bases de données. Il est accessible à partir du lien URL : <http://www.ncbi.nlm.nih.gov/>.

- **UniprotKB**

C'est une banque de protéines regroupant deux groupes: SwissProt (protéines issues d'expérimentations validées et annotées) et TrEMBEL (protéines provenant d'analyses informatiques et non validées expérimentalement. Il est accessible à partir du lien URL : <http://www.uniprot.org/>.

- **Norine**

Norine est une base de données publique, qui a pour but de répertorier l'ensemble des peptides NRPS et des informations liées à leur structure, les monomères les composant et aussi des publications scientifiques se rattachant à ces molécules. Elle a été mise au point par la collaboration entre deux équipes de recherche (ProBioGEM et le Laboratoire d'Informatique Fondamentale de Lille) de l'Université de Lille 1, France. Norine contient un peu plus de 1070 peptides (Caboche *et al.* 2008).

Elle est accessible à partir du lien : <http://bioinfo.lifl.fr/norine/>.

- **Integrated Microbial Genome (IMG)**

C'est un des portails que nous avons consulté pour l'exploration des génomes bactériens séquencés et annotés. Il a été créé par Joint Genome Institute (JGI), Californie, USA.

Il contient un peu plus de 5000 génomes séquencés et annotés, dont environ 500 qui l'ont été partiellement (Markowitz *et al.* 2007).

- **Bioinformatic ToolBox**

C'est un portail bioinformatique qui contient plusieurs outils mais uniquement le logiciel NRPS Predictor a servi lors de cette étude. NRPS Predictor permet la détermination de l'acide aminé activé par les domaines d'adénylation des synthétases (Rausch *et al.* 2005). Il est accessible à partir du lien URL : www-ab.informatik.uni-tuebingen.de/.

Plusieurs autres supports bioinformatiques ont été consultés et seront signalés à chaque fois qu'ils auront été utilisés pour l'analyse de résultats.

Chacun des outils précités nécessite un format particulier des séquences, mais le plus généralement adapté est le format Fasta. Grâce à RedSeq sur <http://bioinfo.hku.hk/>, il est possible de convertir tous les fichiers au format désiré.

II. Méthodes

2.1. Techniques moléculaires

La technique de PCR a été utilisée pour la recherche des gènes NRPS des souches de *Lactobacillus* et de *Bacillus*, elle l'a aussi été pour l'identification moléculaire de certaines souches lors de cette étude. Elle s'effectue sur des ADNs qui sont extraits de souches et de colonies isolées.

2.1.1. Extraction de l'ADN

L'extraction a été réalisée à partir de cultures incubées pendant une nuit à 30°C dans différents milieux notamment dans le milieu MRS pour les souches de *Lactobacillus* et le milieu LB pour les souches de *Bacillus*. L'extraction est réalisée à l'aide du Kit Wizard Genomic DNA en suivant les instructions du fournisseur (Promega Corp., Madison, USA). Elle se déroule généralement en 3 étapes décrites ci-après :

- la lyse des cellules avec du lysozyme (10 mg/mL) et de l'EDTA (50 mM).
- la destruction des protéines et de l'ARN à l'aide d'une solution de protéase et ribonucléase.
- la précipitation par l'éthanol 70% et l'isopropanol, puis l'élution de l'ADN soit dans de l'eau ou dans un tampon approprié selon l'usage prévu.

L'ADN ainsi extrait est quantifié puis conservé à -20°C pour une utilisation future.

La quantification de l'ADN a été réalisée soit à l'aide du NanodropTM 1000 Spectrophotometer (1,5 µL d'extrait d'ADN sont déposés sur la lunette optique de l'appareil puis la concentration est donnée automatiquement), soit une estimation par comparaison avec un marqueur de taille (O' Gene Ruler ou Track-ItTM 100 bp ADN Ladder), acquis respectivement de chez Fermentas et Invitrogen après une migration dans un gel d'agarose.

2.1.2. Préparation des échantillons pour PCR sur colonies

La technique de PCR sur colonies a été mise au point à la fois pour les souches *Lactobacillus*, *Bacillus* et *E. coli*. Elle a été utilisée lors de criblage de collections de souches.

Un isolement sur boîte de Pétri contenant le milieu spécifique de croissance de la souche a été réalisé et incubé une nuit à la température appropriée. Ensuite, une colonie au choix est prélevée et mise dans un tube (de préférence un tube Eppendorf) contenant 25 µL d'eau stérile. L'échantillon est congelé à -20°C pendant 20 min et le tube est directement porté dans un bain-marie bouillant à 100°C pendant 2 min. Le mélange est centrifugé à 8978 g pendant 5 min et 1,5 µL du surnageant est utilisé pour réaliser la PCR.

La variation de température (de -20°C à 100°C) détruit la paroi cellulaire et les insolubles sont éliminés centrifugation.

Pour les vérifications de la présence des inserts, 1 µL des amorces spécifiques du vecteur pJET1/Blunt (pJet FSP et pJet RSP) a été utilisé avec 10 µL de MasterMix pour un volume total de 20 µL de produit PCR. Entre 10-30 µM d'amorces spécifiques dans un volume final de 25 µL ont été utilisés pour le criblage de souches.

2.1.3. Conditions expérimentales de la PCR

Les conditions expérimentales ont été déterminées pour trouver les conditions optimales des composés constitutifs pour la réalisation d'une PCR :

- Détermination de la concentration optimale de l'ADN

La détermination de la concentration optimale de l'ADN a été effectuée en testant plusieurs dilutions d'un ADN de concentration connue.

- Détermination de la température optimale d'hybridation

La température optimale d'hybridation des amorces a été déterminée en réalisant un gradient de température autour d'une température d'hybridation théorique ($\pm 10^\circ\text{C}$). La température d'hybridation théorique (T_h) est déterminée en fonction de la melting température (T_m) qui est généralement fournie par les fournisseurs pour chaque amorce. Cette T_m dépend de la taille de la séquence, du % GC et de la concentration des cations présents dans la solution en particulier de l'ion sodium (Na^+). Elle est déterminée de deux manières différentes:

Lorsque la séquence de l'amorce est comprise entre 14-20 bp (Formule de Wallace)

$$T_m (^{\circ}\text{C}) = 2 \times (A + T) + 4 \times (C + G)$$

Lorsque la séquence de l'amorce est supérieure à 20 nucléotides (Formule empirique)

$$T_m (^{\circ}\text{C}) = 60,31 \times (\%G + \%C) - \left(\frac{675}{N}\right)$$

N = taille de la séquence

La température optimale d'hybridation est obtenue expérimentalement à l'aide de l'appareil MasterCycler PCR Gradient. Elle est généralement égale à :

$$T_h = T_m \pm 5^{\circ}\text{C}$$

Les concentrations initiales des amorces ont été fixées à 20 µM pour toutes les PCRs réalisées.

Dans un tube PCR de volume 0,2 mL ou 0,5 mL, on a mis 1,5 µL de la solution d'ADN (10-30 ng/µL), 1,5 µL d'une solution de chacune des deux amorces (pour les brins direct et inverse) de

concentration initiale de 20 μM , 12,5 μL de la *Taq*-polymerase Master Mix 2x concentré et 8 μL d' H_2O , l'ensemble constituant un volume de 25 μL .

Les programmes pour chacune des PCR ont été principalement définis en fonction de la nature des amorces et de la taille des fragments à amplifier.

2.1.4. Electrophorèse

La migration électrophorétique des produits PCR a été essentiellement réalisée dans des gels d'agarose. Le gel horizontal d'agarose a été préparé dans un tampon TBE 0,5x à pH 8.8 (pour 1 L: 53,9 g de Tris, 27,5 g d'acide borique, 3,65 g d'EDTA acide) additionné de 1% de bromure d'éthidium (BET) à 1 mg/mL. Les échantillons sont additionnés d' $1/10^{\text{ème}}$ de volume de solution de chargement ADN ou bleu de charge (bleu de bromophénol 2,5 g/l, xylène cyanol 2,5 g/l, Ficoll 400 250 g/l).

La migration s'effectue dans le tampon TBE 0,5x sous une différence de potentiel de 50-100 volts à l'aide du Mupid-Ex, Japan. Les molécules d'acides nucléiques sont ensuite visualisées sous UV dans l'Illuminator/Digitally Recorder (GelDOC Bio-Rad, Hercule, USA), grâce à la fluorescence du BET qui s'intercale entre les bases nucléotidiques.

Le gel d'agarose a été préparé à 0,5 % ou 1 % (w/v) selon que les fragments d'ADN à visualiser sont de petite taille (inférieur à 1 Kb) et à 2% (w/v) pour des fragments de grande taille (supérieur à 2 Kb).

Le logiciel Quantity One incorporé au GelDoc permet une annotation et une exportation de l'image du gel sous format TIFF.

2.1.5. Purification des acides nucléiques

Cette opération consiste à débarrasser le fragment d'intérêt des autres produits PCR (amorces, sels, enzymes, dNTP restants). Elle a été effectuée à partir des produits directement issus des PCR ou du fragment après l'électrophorèse.

- A partir de gel d'agarose

Les morceaux de gels contenant les fragments désirés sont coupés à l'aide d'un cutter stérilisé sous flamme pour chaque bande et sont purifiés suivant le protocole QIAquick Gel Extraction Kit (Qiagen, Hilden, Allemagne). Il est indispensable d'exposer le moins longtemps possible les bandes à extraire aux U.V afin d'éviter la dégradation des ADN.

- A partir de produits PCR

Les produits issus de l'amplication PCR ont été purifiés suivant le protocole High Pure PCR (Product Purification Kit, Roche, Allemagne).

2.1.6. Ligation et clonage

Les fragments d'ADN issus d'amplification par PCR sont clonés, après purification, directement dans le vecteur pGEM-T Easy de Promega et après adaptation des extrémités 3'cohésives dans le vecteur pJET1/Blunt de Fermentas selon les instructions des différents fournisseurs.

Les produits de ligation sont laissés une nuit à 16°C en présence d'une unité de T4 DNA ligase, dans le tampon de ligation fourni par le fournisseur correspondant.

La quantité d'ADN à ajouter au mélange de ligation se calcule selon la formule suivante:

$$[\text{ng de vecteur}] \times [\text{rapport de taille (kb) insert/vecteur}] \times [\text{rapport molaire insert/vecteur}]$$

2.1.7. Transformation et sélection des colonies

La transformation a consisté à transférer le vecteur d'intérêt dans une cellule compétente. Alors que les cellules compétentes *E. coli* JM109 ont été acquises commercialement (Promega Corp., USA), les cellules XL1-Blue ont été préparées selon le protocole suivant :

Une préculture d'*E. coli* XL1-B est incubée durant une nuit dans du milieu 853 stérile. Ce milieu estensemencé avec au maximum $1/50^{\text{ème}}$ du volume soit 2 mL pour 100 mL de milieu 853 de la préculture, puis incubé à 37°C pour atteindre une D.O à 600 nm de 0,4. Une centrifugation à froid est réalisée pendant 10 min à 8217 g. Le culot est resuspendu dans la moitié du volume d'ensemencement de CaCl_2 50 mM froid (soit 50 mL) et laissé maintenu dans la glace.

Une nouvelle centrifugation de 10 min à 8217 g est réalisée et le culot obtenu est suspendu dans $1/10^{\text{ème}}$ du volume d'ensemencement soit 9 mL de CaCl_2 50 mM + glycérol 15% et 1 mL de CaCl_2 50 mM froid.

Les cellules compétentes ainsi obtenues sont conservées à - 80°C pour une utilisation ultérieure.

La transformation s'effectue par choc thermique dans le milieu SOC selon les recommandations du fournisseur.

Une partie de l'aliquot d'environ 200 μL de la suspension bactérienne contenant 2 μL du produit de ligation est étalée sur le milieu LB solide complémenté de $100 \mu\text{g.mL}^{-1}$ d'ampicilline, de X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside), puis de l'IPTG (isopropyl- β -D-thiogalactopyranoside). Ces composés sont ajoutés stérilement après une filtration sur membrane de 0,45 μm de porosité à une concentration de $200 \mu\text{g.mL}^{-1}$ dans le cas d'utilisation du vecteur pGEM-T Easy. La sélection des transformants est effectuée selon le test dit « blanc-bleu ».

En effet, X-Gal et IPTG permettent de détecter les bactéries contenant le vecteur avec l'insert suite à une réaction entre X-Gal et le gène de la β -galactosidase présent dans le vecteur et induit par l'IPTG:

Substrat incolore chromogène X-Gal $\xrightarrow{\beta\text{-galactosidase}}$ Galactose + X (produit coloré bleu).

Pour les colonies possédant le vecteur et l'insert, le gène de la β -galactosidase sera rendu inactif par l'insertion du fragment d'ADN, ce qui conduira à des colonies blanches.

Les colonies bleues auront une β -galactosidase fonctionnelle et ne posséderont donc pas l'insert.

Lorsque le vecteur pJET1/Blunt a été utilisé, 200 μ L de la suspension bactérienne a été étalé sur le milieu 856 complémenté d'ampicilline (100 μ g.mL⁻¹). Les transformants sont sélectionnés après une nuit d'incubation à 37°C car l'endonucléase létale empêche la croissance de celles dépourvues d'inserts.

Pour éviter le choix de « faux positifs » ou mutants, des vérifications autres que ces caractères biochimiques ont été réalisées. En effet, deux techniques ont été utilisées pour vérifier la présence du fragment inséré dans le vecteur :

- soit la digestion par l'enzyme de restriction *EcoRI* (Fermentas, France) dans le cas du vecteur pGEMT-Easy dans les conditions suivantes: 1 μ L d'enzyme de restriction *EcoRI* (10u/ μ L), 1 μ L tampon *EcoRI* et 8 μ L de l'ADN plasmidique concentré à 30-60ng/ μ L. Le mélange est incubé à 37°C pendant 1 h 30 min.
- soit par PCR directement sur colonies comme décrit précédemment avec 1 μ L des amorces pJet FSP/pJet RSP dans le cas du vecteur pJET1/Blunt.

2.1.8. Extraction de plasmides

L'ADN plasmidique a été extrait des cellules transformées d'*E. coli* à l'aide de QIAprep Spin Miniprep Kit Protocol (Qiagen, Allemagne) après culture d'une nuit dans le milieu 853 ou LB complémenté en ampicilline (concentration finale à 100 μ g/mL).

2.1.9. Séquençage

Le séquençage des fragments clonés a été effectué à la fois chez Cogenics S.A (Meylan, France) et MWG Eurofins, Martinsried (Allemagne). Les échantillons ont été préparés selon les recommandations de l'entreprise concernée. Les produits clonés sont séquencés en utilisant les amorces pUC-M13-R/F dans le système ABI PRISM DYE Terminator cycle sequencing kit.

2.2. Techniques bioinformatiques

Des techniques bioinformatiques ont été utilisées pour la définition des amorces spécifiques. En effet, les amorces constituent un élément clef pour la détection de gènes par PCR.

Ces techniques ont consisté principalement à l'utilisation des logiciels comme Vector NTI Advance 10TM et AmplifX 1.4.4 et des liens Webs cités précédemment.

2.2.1. Détermination des amorces à partir des gènes NRPS de *L. plantarum*

L'objectif a été de définir des amorces capables d'amplifier des fragments de gènes codant pour des synthétases non ribosomiales.

Les amorces pouvant être utilisées par PCR ont alors été conçues à partir des gènes NRPS (*npsA* et *npsB*) du génome de *L. plantarum* WCSF1 (Kleerebezem *et al.* 2003). Ce génome a été exploré à l'aide du logiciel Vector NTI connaissant son numéro d'accension Genbank: AL935263.

- Les amorces non dégénérées

Ce type d'amorces est composé uniquement de combinaisons d'acides nucléiques A, T, G et C qui forment une séquence unique et n'est susceptible d'amplifier qu'un seul fragment. Deux couples d'amorces non dégénérées (A7.1/T1.1 et A10.4/C3.4) ont été définies comme explicité dans le Chapitre 5 suivant.

- Amorces dégénérées

Les amorces dégénérées comportent dans leurs séquences des bases nucléiques capables de substituer 2, 3 voire 4 autres bases nucléiques. On parle de dégénérescence car la séquence littérale est en réalité la résultante de plusieurs séquences, ce qui permet de reconnaître plusieurs fragments.

Le code DNA IUPAC (*International Union of Pure and Applied Chemistry*) a été utilisé pour introduire ces dégénérescences dans les séquences des amorces.

Des alignements des domaines d'adénylation et de thiolation du gène *npsA* ont été réalisés à l'aide du logiciel Vector NTI et, selon les séquences consensus nucléiques obtenues, deux couples d'amorces ont été définis :

- le couple d'amorce A7.d1/A10.d1 a été obtenu à partir des alignements de séquences nucléiques des 6 modules composant les gènes *npsA* et *npsB*.

- le couple Alb1-F/Tlb1-R a été déterminé à partir des alignements des séquences nucléiques des domaines d'adénylation et de thiolation composant le gène *npsA*.

Les motifs les plus conservés de ces alignements sont choisis à l'aide de Vector NTI qui permet de déterminer en pourcentage la position la plus consensuelle des nucléotides formant un motif. Le choix des motifs conservés nucléiques utilisés pour ces amorces s'est fait selon les critères ci-après :

- le degré de consensus exprimé en pourcentage qui dépend du taux de similarité
- le sens d'orientation des amorces : sens (5'3') ou non sens (3'5').

- le niveau de dégénérescence qui tient compte de la stabilité de la sonde par détermination du coefficient d'hybridation.

Ainsi, deux couples respectivement d'amorces non dégénérées et dégénérées ont été définis à partir des gènes NRPS de *Lactobacillus plantarum* WCSF1 pour la recherche de molécules non ribosomiales.

2.2.2. Détermination des amorces pour la détection de gène NRPS de *Bacillus*

La définition des amorces spécifiques tient compte de la grande diversité des opérons impliqués dans la synthèse de molécules non ribosomiales. Les bacilles sont connus pour leurs synthèses de diverses molécules NRPS.

Des séquences nucléiques des domaines A et T des opérons de synthétases impliqués dans la synthèse de lipopeptides de *Bacillus* ont servi pour définir les amorces selon des critères ci-après:

- Sélectivité

Les amorces doivent amplifier uniquement des gènes de type NRPS, c'est-à-dire que tous les fragments détectés doivent appartenir à des gènes codant uniquement pour des synthétases de type NRPS. C'est pourquoi la région formée par les domaines A et T exclusifs aux NRPS selon Schwarzer *et al.* (2001) seront utilisés.

- Spécificité

Les amorces doivent amplifier des fragments de gènes pour lesquels elles sont définies en priorité. Le cas échéant, en absence du fragment recherché, elles pourraient amplifier d'autres gènes de types NRPS en particulier ceux impliqués dans la synthèse des gènes dont les synthétases sont proches du gène ayant servi à leur détermination.

- Efficacité

Les gènes amplifiés par ces amorces doivent avoir une forte homologie avec des gènes NRPS existant. Cette efficacité des amorces est évaluée par l'expect value ou E. value des séquences amplifiées par les amorces. La E. value est la probabilité d'avoir un meilleur résultat que celui obtenu après une analyse Blast (*Basic Local Alignment Search Tools*). En effet, plus la E. value est proche de 0.00, plus le pourcentage d'identité du fragment affiché est bon, ce qui atteste de l'efficacité des amorces. Certains définissent la « E. value » comme un indice de confiance, une valeur probabilistique inverse.

Cinq couples d'amorces ont été ainsi déterminés : les amorces As1-F/Tf1-R et Ap1-F/Tp1-R respectivement à partir des opérons de la surfactine et de la plipastatine de *B. subtilis* 168, les amorces Af2-F/Tf1-R et Am1-F/Tm1-R respectivement à partir des opérons de la fengycine de

B. subtilis F29-3 et de la mycosubtiline de *B. subtilis* ATCC 6633 enfin les amorces Abl1-F/Tbl1-R à partir de la bacillomycine de *B. subtilis* AU195.

Le choix des motifs conservés nucléiques utilisés pour ces amorces s'est fait selon les critères définis dans le cas des amorces lactobacilles.

Puisque ces amorces devront être capables de détecter des fragments de gènes appartenant aux opérons ayant servi à leur définition mais aussi aux autres opérons des lipopeptides de la même famille protéique, une étude sur la similarité des opérons de molécules issues de la même famille a été réalisée.

- Conditions de comparaison des synthétases

L'étude de la comparaison des opérons a été réalisée par des alignements multiples à l'aide de EMBOSS (Needle, Global) Pairwise Alignment Algorithms (Needleman and Wunsch 1970), disponible sur <http://www.ebi.ac.uk>. L'alignement des séquences s'est fait dans les conditions suivantes : Protéines : Matrix : Blosum62, Open Gap: 10, Extend Gap: 0.5. ADN: Matrix : ADN full, Open Gap: 10, Extend Gap: 0.5.

Les synthétases des principaux lipopeptides qui constituent chacune des 3 familles ont ainsi été comparées entre elles au niveau protéique et nucléique.

Les séquences des différentes synthétases ont été extraites dans la banque Genbank de NCBI avec les numéros d'accèsion : surfactine (CAB12142.1, CAB12143.1 et CAB12145.1), plipastatine (CAB13717.1, CAB13716.1, CAB13715.1, CAB13713.1 et CAB13714.1), mycosubtiline (AAF08795.1, AAF08796 et AAF08797), fengycine (AAB80955, AAB00093, CAA09819 et AAB80956.1) et bacillomycine (ABS 74181.1, ABS74180.1 et ABS74179.1).

La reconnaissance des séquences nucléotidiques des modules et des domaines correspondant à chaque synthétase a été réalisée à l'aide de Vector NTI Advance 10TM, en complément du logiciel NRPS-ProBioGEM.

- Vérification des amorces

Pour déterminer la stabilité des amorces, un Coefficient d'Hybridation (HyC) a été mis au point par une approche mathématique rendant compte du niveau de dégénérescence de l'amorce (Chapitre 5). A chaque acide nucléotidique dégénéré selon le code IUPAC, un coefficient a été affecté comme présenté dans le Tableau 4.

Le coefficient affecté est fonction du nombre de nucléotides que peut substituer chaque base dégénérée.

Tableau 4. Coefficients affectés à chaque acide nucléique dégénéré

Acide nucléique dégénéré	Bases substituables	Coefficient affecté
R	A ou G	2
Y	C ou T	2
M	C ou A	2
K	T ou G	2
W	T ou A	2
S	C ou G	2
B	C, T ou G	3
D	A, T ou G	3
H	A, T ou C	3
V	A, C ou G	3

La validation des amorces a nécessité l'utilisation du logiciel AmplifX 1.4.4, à l'aide duquel des tests PCR ont été réalisés. On parlera de tests *in silico*. Ce logiciel a permis de déterminer plusieurs autres caractéristiques de ces amorces, notamment leur position sur les opérons, l'hybridabilité ou la capacité à amorcer et surtout la taille théorique des fragments à amplifier.

Les amorces ainsi définies ont été testées *in vitro* avec les souches de références pour déterminer les conditions optimales de températures pour la réalisation des tests de PCR.

Elles ont été ensuite utilisées *in vitro* sur les souches étudiées pour la recherche de gènes de type NRPS.

- Conditions expérimentales d'utilisation des amorces par PCR

Les conditions générales pour les PCR ont été effectuées pendant 30 cycles avec une température d'extension finale de 72°C pendant 10 min.

Les conditions spécifiques de températures optimales des étapes principales (dénaturation, hybridation et élongation) pour chacune des amorces sont présentées dans le tableau 5.

Tableau 5. Conditions expérimentales de programmation de l'appareil PCR pour l'utilisation des amorces

Étapes	Amorces <i>Lactobacillus</i>	Amorces <i>Bacillus</i>
Dénaturation	94°C pendant 4 min pour toutes les amorces (étape initiale) puis pendant 1 min pour les autres cycles	94°C pendant 3 min pour toutes les amorces (étape initiale) puis pendant 1 min pour les autres cycles
Hybridation	45°C pendant 30 s (A7.d1/A10.d1, Alb1/Tlb1-R, A7.1/T1.1 et A10.4/C3.4)	43°C pendant 30 s pour As1-F/Ts2-R, 45°C pendant 30 s pour Am1-F/Tm1-R, Af2-F/Tf1-R et Abl1-F/Tbl1-R 58°C pendant 30 s pour Ap1-F/Tp1-R
Elongation	72°C pendant 45 s (A7.d1/A10.d1, Alb1/Tlb1-R, A7.1/T1.1 et A10.4/C3.4)	72°C pendant 45 s pour As1-F/Ts2-R, Am1-F/Tm1-R, Af2-F/Tf1-R et Abl1-F/Tbl1-R 72°C pendant 75 s pour Ap1-F/Tp1-R

Les amorces spécifiques PrBac1 et Prbac2 (Rupf *et al.* 1999) ont été utilisées comme témoin-contrôle de la qualité des PCR dans les conditions expérimentales de chaque amorce utilisée.

2.3. Identification des souches

L'identification des souches a été réalisée de deux manières : identification moléculaire et biochimique.

2.3.1. Identification moléculaire

- Condition d'utilisation des amorces pour l'identification moléculaire

Des amorces ont été utilisées *in vitro* pour l'identification des souches isolées des condiments fermentés africains (Tableau 6).

Tableau 6. Caractéristiques des amorces utilisées pour l'identification des souches isolées

Amorces	Séquences	Fragments attendus	Genres et espèces concernés
B-K1/F B-K1/R1	TCACCAAGGCRACGATGCG CGTATTCACCGCGGCATG	1114 bp	<i>Bacillus</i> , <i>Paenibacillus</i> , <i>Brevibacillus</i>
Bsub5F Bsub3R	AAGTCGAGCGGACAGATGG CCAGTTTCCAATGACCCCTCCCC	595 bp	Groupe « <i>Bacillus subtilis</i> » (<i>subtilis</i> , <i>licheniformis</i> , <i>amyloliquefaciens</i> , <i>pumilus</i> , <i>atrophaeus</i>)
Ba1F Ba1R	TGCAACTGTATTAGCACAAGCT TACCACGAAGTTTGTTCACTACT	533 bp	Groupe « <i>Bacillus cereus</i> » (<i>cereus</i> , <i>thuringiensis</i> , <i>mycoides</i> , <i>anthracis</i>)

Wu *et al.* (2005) ont défini les amorces K-B1F/K-B1R1 capables d'amplifier le gène de l'ARN 16S de souches appartenant au genre *Bacillus*, *Paenibacillus*, *Brevibacillus*.

Les amorces Bsub 5F/Busb 3R, mises au point par Wattiau *et al.* (2001) sont spécifiques pour la détection du gène de l'ARN 16S, permettant ainsi d'identifier 5 espèces du groupe « *Bacillus subtilis* » (*B. subtilis*, *B. pumilus*, *B. atrophaeus*, *B. licheniformis* et *B. amyloliquefaciens*). Chang *et al.* (2003) ont déterminé les amorces Ba1F/Ba1R pour l'identification du gène *groEL* présent chez le groupe « *B. cereus* ».

Ces amorces ont été utilisées pour caractériser les souches issues des collections de lactobacilles et de bacilles.

2.3.2. Identification biochimique

Les galeries API 50 CHB (Biomérieux SA, Marcy-l'Etoile, France) ont été utilisées et inoculées selon les recommandations du fournisseur. L'incubation est réalisée à 30°C et la lecture est effectuée après 24 et 48 h d'incubation.

Le logiciel APIWEB a été utilisé pour le traitement des données en vue de déterminer l'espèce.

2.3.3. Détermination d'activités biologiques

Les activités biologiques ont été déterminées pour des souches de *Lactobacillus* et de *Bacillus*.

- Tests antifongiques

Deux techniques (double couche et puits) ont été effectuées pour le test antifongique. Elles consistent à observer la capacité de la souche testée à provoquer une inhibition de différentes levures et de spores de moisissures.

Les spores de moisissures ont été préparées de la manière suivante :

Quelques μL de moisissures ont été déposées sur une gélose PDA 17 g.L^{-1} additionnée de chloramphénicol ($0,5 \text{ g.L}^{-1}$). Ces boîtes ont été incubées à 20°C pendant une dizaine de jours.

4,5 mL d'eau peptonée contenant du tween à 0,1% stérile sont répartis à la surface de la boîte de Pétri contenant la moisissure. La surface de la gélose est alors raclée à l'aide d'un râteau afin de récupérer les spores. Après filtration sur fibres de verre stériles, la concentration en spores de la suspension est déterminée par numération en cellule de Malassez.

La détermination des spores est calculée à l'aide de la formule suivante :

$$N = \left[\frac{n \cdot uT}{u \cdot V} \right] \cdot FD$$

N : nombre spores par mL ; u : nombre d'unités comptées ; uT : nombre total d'unité de comptage

n : nombre de cellules comptées V : volume de l'hématimètre. FD: facteur de dilution.

Dans la cellule de Malassez, une unité correspond à un grand carré soit à un champ optique.

- La technique de la double couche

Elle consiste à inoculer une gélose MRS à 1,7 % d'agar (Magnusson *et al.* 2002) par une strie très fine de suspension bactérienne à l'aide d'une pipette Pasteur. Pour les lactobacilles, le milieu est ensuite incubé pendant 48 h sous atmosphère enrichie en CO_2 .

Après incubation, les boîtes de Pétri sont séchées à l'étuve pendant 1 h, inoculées en double couche par 4 mL de gélose PDA à 0,7 % contenant 10^4 spores de moisissure par mL.

Les boîtes sont ensuite ré-incubées durant 24 à 48 h à température ambiante ou dans une étuve à 20°C . La lecture s'effectue en mesurant la largeur du halo d'inhibition présent autour de la strie.

- **La technique des puits**

Elle consiste à inoculer des boîtes de Pétri contenant 20 mL de milieu PDA à 1,5 % d'agar par 4 mL de gélose à 0,7 % d'agar contenant 10^4 spores par mL de milieu. Après séchage des boîtes à l'étuve (37°C pendant 1 h), 4 puits sont réalisés dans le milieu à l'aide de tube à hémolyse stérile. 100 μL de surnageant de culture de culture concentré 10 fois sont déposés dans les puits. Un contrôle (puits vide) et un témoin négatif, contenant le milieu non inoculé et concentré au même titre que le surnageant de culture, sont réalisés. Les boîtes sont ensuite incubées à 30°C pendant 24 à 48 h sous atmosphère enrichie en CO_2 .

La lecture s'effectue en mesurant le diamètre du halo d'inhibition présent autour du puits. L'inhibition est traduite par un score : - : pas d'inhibition, + : zone d'inhibition entre 0,1 et 0,5 cm, ++ : zone d'inhibition entre 0,5 et 1 cm, +++ : zone d'inhibition supérieure à 1 cm.

- **Test de compétitions cellulaires**

Le test dit « de compétition cellulaire » consiste à inoculer sur une boîte gélosée contenant un milieu d'incubation approprié, une colonie de souche dont on veut déterminer l'activité antifongique en présence de 2 μL de spores de moisissure sur une distance de 2 cm. L'incubation se fait pendant 24 h à température optimale de croissance de la souche. La lecture s'effectue par observation d'une inhibition de croissance de la moisissure par la colonie.

- **Test hémolytique**

L'hémolyse, ou lyse des hématies, est due à la rupture de leur membrane plasmique, souvent sous l'action de phosphatidyl-choline estérase des bactéries et sous l'action de molécules provoquant la formation de pores membranaires (hémolysine).

Ce phénomène libère de l'hémoglobine qui est ensuite plus ou moins digérée. Si la digestion est totale, la couleur rouge disparaît et on observe une zone éclaircie (hémolyse partielle), voire incolore (hémolyse complète) autour de la colonie. On parle alors d'hémolyse β . La digestion peut être incomplète s'il se forme des produits verdâtres ou marron (dont la méthémoglobine), on parle alors d'hémolyse α .

On distingue donc :

- Les germes à hémolyse β complète et hémolyse β partielle
- Les germes à hémolyse α
- Les germes non hémolytiques (parfois appelée hémolyse ϕ)

Dans 100 mL de milieu LB gélosé, 5 mL de sang de cheval y sont ajoutés. Ces boîtes sontensemencées par repiquage de 100 μL de culture de souches puis incubées à la température

optimale de croissance de la souche pendant 24 h. Les zones d'inhibition ou auréoles sont observées.

2.3.4. Caractérisation des souches isolées des condiments fermentés africains

- Echantillonnage

• *Soumbala*

Le Soumbala est un condiment fermenté fait à partir des graines du fruit de *Parkia biglobos*.

La figure 18 présente une photographie de Soumbala en boulettes vendues sur les marchés du Burkina-Faso.



Figure 18. Photographie de soumbala

Il est présent sur la quasi-totalité des marchés ouest-africains. Les ménagères s'en servent pour la confection de mets comme condiments pour l'assaisonnement des sauces locales.

On le consomme de diverses façons: soit sous forme de sauce (les graines sont grillées, brisées et fermentées dans de l'eau pour sa confection), soit sous forme pelée (une pâte obtenue par broyage des graines après fermentation).

Des échantillons ont été récupérés dans différents endroits du Burkina-Faso (Peni, Bobo-Dioulasso, Yaar de 1200 Ouaga, Toussiana et de Sala).

• *Bikalga*

Le Bikalga est un condiment fermenté fait à partir des graines d'oseille (*Hibiscus sabdarifa*). Ce condiment est connu comme contenant plusieurs souches de *Bacillus* jouant un rôle lors de la fermentation pour la fabrication du produit (Parkouda *et al.* 2008).

Cinq échantillons de Bikalga (Yaar 1200, Ouaga A, Ouaga B, Ouaga C, Ouaga D) ont été utilisés.

Les souches ont été isolées à l'aide du milieu MRS pour les bactéries lactiques, de la gélose nutritive pour la flore aérobie totale. Les différentes dilutions et solutions mères ont été

effectuées dans de l'eau peptonée. L'identification a été effectuée à l'aide des tests biochimiques et moléculaires.

Pour éviter les redondances, une sélection a été effectuée sur la base des caractères morphologiques des colonies (forme, couleur et taille), sur la capacité des colonies à dismuter le H_2O_2 (test catalase) et sur leurs diverses activités biologiques. Le genre *Bacillus* responsable d'activité biologique a été retenu pour la recherche de lipopeptides.

- Codification des souches

Des codes spécifiques ont été attribués à chaque colonie isolée en fonction de sa température optimale de croissance, du milieu d'incubation, de l'échantillonnage et du type d'aliment.

Exemple : **0XaB0YZ** signifie que la souche a été isolée à 30°C, **X** correspond au type d'aliment (**S** pour Soumbala et **B** pour Bikalga), **a** au numéro de l'échantillon, **0Y** au numéro de la colonie lors du premier isolement et **Z** lorsqu'elle a été réisolée.

2.4. Recherche de lipopeptides non ribosomiaux

2.4.1. Mesure de la tension de surface

La présence de molécules biosurfactantes telles que les lipopeptides est mise en évidence par la mesure de la tension de surface des surnageants de culture. Elle permet de déterminer la teneur relative du milieu en biosurfactant: plus la mesure est élevée, plus le milieu est pauvre en biosurfactant. Au contraire, plus le milieu est riche en biosurfactant, plus la tension de surface est faible, jusqu'à une Concentration Micellaire Critique (CMC). Il peut donc être nécessaire de diluer l'échantillon afin d'être en dessous de la CMC.

Les cultures bactériennes incubées de 24 h à 120 h à la température optimale de croissance des souches utilisées sont centrifugées (15000 g pendant 5 min). La mesure s'effectue selon la méthode de Du Nouy à l'aide du tensiomètre (Lauda, Königshofen, Allemagne) sur un volume de 20 ml du surnageant placé dans une cupule en verre préalablement traitée à la potasse alcoolique. Elle s'exprime en $mN.m^{-1}$.

L'interprétation des résultats se fait par rapport à la tension de surface mesurée sur le milieu de culture. Des valeurs inférieures à cette dernière indiquent la présence de biosurfactants.

2.4.2. Recherche de lipopeptides par HPLC

La technique chromatographique HPLC est basée sur la différence d'affinité des molécules vis-à-vis de la phase stationnaire hydrophobe et de la phase mobile hydrophile. Elle a été utilisée pour la recherche de lipopeptides dans les surnageants de culture.

Le détecteur est réglé à des longueurs d'onde de 215 nm pour la détection des liaisons peptidiques. Une colonne C18 VYDAC 218 TP de dimensions 4,6 × 250 mm (Hesperia, CA) est utilisée à un débit de 1 mL.min⁻¹.

Les solvants utilisés sont le mélange acétonitrile-eau-acide trifluoroacétique (TFA) : 40/60/0,5 v/v/v et 80/20/0,5 v/v/v respectivement pour les fengycines, les iturines et les surfactines.

Les échantillons de lipopeptides ont été obtenus à partir des surnageants purifiés comme décrit ci-après :

La souche a été incubée pendant 24 h à 30°C dans le milieu de culture approprié. 1 mL de surnageant récupéré après centrifugation (15000 g pendant 10 min à 5°C) est chargé sur une colonne C18 900 mg Altech (Grace-Darison, Templemars, France) puis purifié à l'aide d'un dispositif constitué d'une cartouche, de la colonne C18 et d'un manomètre (Savant Instrument, New York, USA) mis sous faible pression selon le procédé suivant.

La colonne est d'abord conditionnée par 20 mL de méthanol 100%, puis rincée par 8 mL d'eau ultrapure. 1 mL de surnageant est ensuite déposé sur la cartouche et rincé par 8 mL d'eau ultrapure puis avec 8 mL de méthanol 50%. La cartouche est séchée pendant 2 min puis l'élution des lipopeptides s'effectue avec 8 mL de méthanol 100%.

Le filtrat est enfin recueilli dans un tube à hémolyse, puis évaporé à l'aide du SpeedVac qui permet un séchage des échantillons par un procédé de sublimation.

Les échantillons ainsi séchés sont repris dans 200 µL de méthanol 100%, et 10 µl de chaque échantillon ont été injectés pour être analysés.

2.4.3. Recherche de lipopeptides par Maldi-TOF/LC-MS/MS-MS

Des échantillons ont été analysés par spectrométrie de masse à l'aide de l'appareil Bruker Ultraflex tof (Bruker Daltonics) selon la méthode décrite par Leenders *et al.* (1999) et Vater *et al.* (2002). C'est une méthode analytique qui permet de déterminer la masse moléculaire des lipopeptides de *Bacillus* à partir d'une colonieensemencée sur milieu solide et à partir de surnageants de culture.

Ces tests ont été réalisés dans le Laboratoire Chimie Biologie du Pr. Whatelet, Gembloux AgroBioTech en Belgique.

Les échantillons ont été préparés soit à partir de surnageants de culture filtrés sur colonne C18 Altech comme décrit précédemment, soit à partir de géloses nutritives. Dans ce dernier cas, la gélose des boîtes ensemencées et incubées pendant 24 h, 48 h, 72 h et 96 h à 30°C est broyée dans de l'eau physiologique (9 g/L NaCl) à l'aide de l'homogénéiseur (Equipements Industriels,

Paris, France). Après centrifugation (15000 g pendant 10 min, répété 3 fois), le surnageant est filtré sur une membrane de diamètre 0,2 µm de pores en vue de réaliser les analyses.

Les échantillons ainsi obtenus sont dilués 10 fois dans une solution saturée d'acide α -cyano-4-hydroxy-cinnamique (3:1 v/v de CH₃CN et H₂O contenant 0.1% d'acide trifluoroacétique).

0.5 µL de cette solution est déposée sur la lunette d'analyse puis la mesure s'effectue en mode reflectron avec une tension de l'analyseur fixé à 20 kV.

PARTIE III.

RESULTATS ET DISCUSSION

Chapitre 5 Définition d'amorces spécifiques pour l'amplification de gènes non ribosomiaux par PCR à partir des gènes NRPS de *Lactobacillus*

Le premier objectif a été de déterminer des amorces capables de détecter des gènes codant pour des synthétases peptidiques non ribosomiales dans des lactobacilles. En effet, la présence de ce type de gène a été mise en évidence pour la première fois en 2003 dans le genre *Lactobacillus* lors du séquençage du génome de l'espèce *plantarum* WCSF1. Par ailleurs, le laboratoire ProBioGEM possède une collection de souches de bactéries lactiques et il était intéressant de vérifier si ces souches avaient un potentiel de synthèse de molécules peptidiques d'origine non-ribosomiale. Dans un premier temps, tous les génomes des lactobacilles séquencés partiellement et/ou complètement, répertoriés dans la banque IMG (Integrated Microbial Genomes) ont été explorés afin de vérifier la présence de gènes susceptibles d'être impliqués dans le mécanisme NRPS. Seule la présence des gènes NRPS, *npsA* et *npsB*, a été montrée dans le génome de *Lactobacillus plantarum* WCSF1 parmi les 17 génomes séquencés de lactobacilles.

Des amorces ont donc été élaborées à partir de ces gènes dans le but de rechercher des gènes NRPS dans la collection des souches de *Lactobacillus* du laboratoire après que celles-ci aient été validées sur des souches modèles.

I. Elaboration des amorces à partir des gènes NRPS de *Lactobacillus*

1.1. Organisation des gènes NRPS de *Lactobacillus*

Les gènes *npsA* et *npsB* codent pour des NRPS qui seraient impliquées dans la synthèse de molécules non ribosomiales. Le gène *npsA* code pour la synthétase NpsA (numéro d'accension : CAD63192.1, 5289 aa) composée de 5 modules, *npsB* code pour la synthétase NpsB (numéro d'accension : CAD63194, 330 aa) formant un module unique.

A ce jour, aucun auteur n'a encore montré de façon expérimentale la nature de la molécule synthétisée par les synthétases NpsA et NpsB de la souche WCSF1, mais les outils de prédiction (NRPS Predictor et NRPS/PKS) permettent par la détermination des acides aminés potentiellement activés, de donner des informations sur sa structure globale.

La figure 19 présente l'organisation modulaire de ces synthétases NpsA et NpsB et des acides aminés activés obtenus par l'utilisation d'outils bioinformatiques.

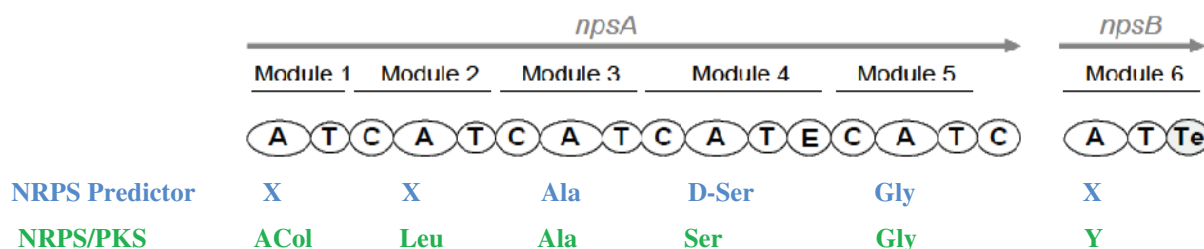


Figure 19. Organisation des modules et domaines des synthétases non ribosomiales de *Lactobacillus plantarum* WCSF1, acides aminés activés par les gènes *npsA* et *npsB*, X : Substrat activé non reconnu; Y : module non reconnu comme étant de type non ribosomal, ACol : monomère non identifié

Il reste toutefois difficile malgré ces prédictions de déterminer la nature et la fonction de la molécule car plusieurs acides aminés restent non déterminés par les logiciels utilisés.

1.1.1. La définition des amorces

Deux types d'amorces ont été déterminés à partir des séquences nucléiques des synthétases NpsA et NpsB.

- Extraction des séquences nucléiques

A partir des numéros d'accèsion CAD63192.1 et CAD63194, les séquences protéiques ont été extraites de la banque Genbank de NCBI et les CDS (CoDing Sequence) ont été récupérées à l'aide du logiciel commercial Vector NTI Advance 10TM. Ce logiciel est en effet capable d'extraire de la banque Genbank la séquence de la protéine et, il permet l'affichage de ces synthétases avec les différentes régions correspondant aux domaines principaux non ribosomiaux.

- Choix des amorces

• Amorces non-dégénérées

Cinq couples d'amorces ont été déterminés à l'aide du logiciel AmplifX 1.4.4, sur la base qu'elles hybrident le gène *npsA*. Les couples permettant d'obtenir des amplicons de taille inférieure à 1000 bp, et qui ne s'auto-hybrident pas ont été retenus. Ainsi, deux couples parmi les cinq, nommés A7.1/T1.1 et A10.4/C3.4 ont été retenus. L'étude comparative des correspondants protéiques montre que les amorces sens A7 et A10 proviennent des séquences consensus d'adénylation respectivement des modules 1 et 4, les amorces reverses T1.1 et C3.4 sont issues

des séquences consensus protéiques respectivement du domaine de thiolation du module 1 et de condensation du module 4 de NpsA (Tableau 1).

- *Amorces dégénérées*

Amorces A7.d1/A10.d1

Un couple d'amorces dégénérées a été défini à partir des alignements de séquences nucléiques des modules des synthétases NpsA et NpsB. On les a nommé A7.d1/A10.d1.

Les 5 modules de NpsA et le module unique de NpsB ont été alignés à l'aide du logiciel Vector NTI. Ensuite à partir des motifs les plus conservés issus de ces alignements, le couple d'amorces a été obtenu comme décrit dans le § 2.2.1 du Chapitre Matériels et Méthodes. La figure 20 présente les alignements réalisés des 6 modules (5 de NpsA et 1 de NpsB) et les motifs conservés qui ont abouti à l'obtention du couple dégénéré A7.d1/A10.d1.

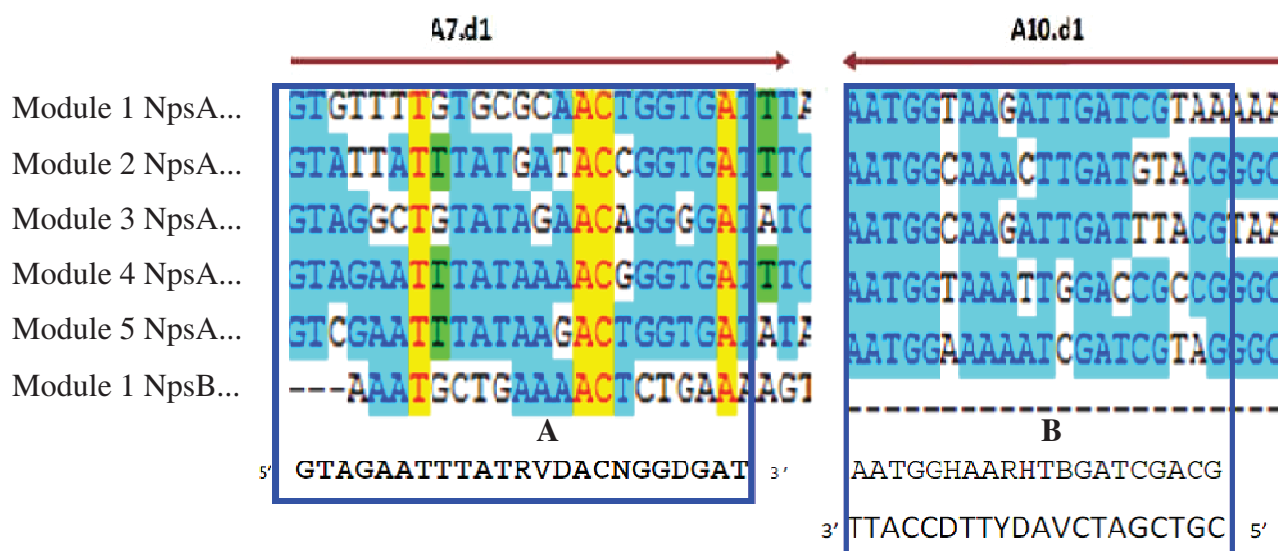


Figure 20. Représentation schématique des alignements de séquences nucléiques et des motifs conservés obtenus par le logiciel Vector NTI (**A** : motif ayant servi à définir A7.d1, amorce sens, **B** : motif ayant servi à définir A10.d1, amorce non sens), les séquences des amorces étant dans le sens 5' 3'

Les positions respectives des deux amorces par rapport aux signatures protéiques ont montré qu'elles se situent toutes les deux dans des domaines d'adénylation des modules alignés. En effet, le motif ayant servi à la définition de l'amorce sens A7.d1 (Figure 20 A) se situe au niveau du motif conservé protéique A7 (voir Tableau 1) alors que le motif ayant servi à la définition de l'amorce non sens A10.d1 (Figure 20 B) se situe au niveau du motif conservé protéique A10 (voir Tableau 1) d'où leurs noms respectifs A7.d1 et A10.d1.

Amorces Alb1-F/Tlb1-R

Parallèlement aux amorces précédentes, un autre couple a été défini à partir des alignements de séquences nucléiques des domaines A et T du gène *npsA*. Le gène *npsB* a été exclu des alignements car ces domaines n'ont pas été reconnus par les logiciels utilisés (figure 19).

Les domaines A et T du gène *npsA* ont été extraits et alignés à l'aide du logiciel Vector NTI. A partir des motifs conservés comme montré à la figure 20 précédente, les amorces sens Alb1-F ont été définies à partir des domaines d'adénylation et les amorces inverses Tlb1-R à partir des domaines de thiolation.

Le tableau 7 présente toutes les amorces obtenues à partir des gènes NRPS de *L. plantarum* WCSF1 et leurs caractéristiques après avoir été testées *in silico* à l'aide du logiciel AmplifX 1.4.4. Cet outil a permis de déterminer la taille et le nombre des amplicons.

Deux couples d'amorces non dégénérées (A7.1/T1.1 et A10.4/C3.4) et deux couples d'amorces dégénérés (A7.d1/A10.d1 et Alb1-F/Tlb1-R) ont ainsi été retenus.

Tableau 7. Caractéristiques des amorces spécifiques NRPS *Lactobacillus* (Vector NTI et AmplifX 1.4.4)

	Nom des amorces	Séquence des amorces	*Position sur <i>npsA</i> et <i>npsB</i>	Taille des amplicons attendus (bp)
Amorces non dégénérées	A7.1 T1.1	TGCGCAACTGGTGATTTAGT ATTCTTTCAATTGGGTGCT	1063-1082 1511-1530	468
	A10.4 C3.4	ATTGGTAAATTGGACCGC GTCATCATCTGATAATTGATAC	10119-10140 10799-10820	680
Amorces dégénérées	A7.d1 A10.d1	GTAGAATTTATRVDACNGGDGAT CGTCGATCVADYTTDCCATT	3663-3686 3393-3413	334, 331, 310, 331, 343, 334
	Alb1-F Tlb1-R	TYTYTGTTGGCACTATTMTTCCA TGAYCYGSTAGCGACWTMCTT	1046-1067 37-57	791, 981

*La position des amorces non dégénérées a été obtenue par le logiciel AmplifX 1.4.4 alors que celle des amorces dégénérées a été obtenue par le logiciel Vector NTI.

Les amorces non dégénérées amplifient un unique amplicon sur *npsA* tandis que les amorces dégénérées en détectent plusieurs, à la fois sur les gènes *npsA* et *npsB*.

De tous les couples définis à partir des séquences consensus, seul le couple A7.d1/A10.d1 a permis d'amplifier *in silico* un fragment dans chacun des 6 modules que constituent les synthétases NpsA et NpsB.

1.1.2. Calcul du Coefficient d'Hybridation

Une valeur mathématique qui rend compte du niveau de dégénérescence des amorces a été mis au point. Elle a été appelée Coefficient d'Hybridation (HyC), et s'exprime sans unités. Ce coefficient a été déterminé pour chacune des amorces composant un même couple pour la PCR. Du niveau de dégénérescence des amorces dépendra la valeur du HyC, qui renseigne sur la stabilité théorique de celles-ci.

Selon le code DNA-IUPAC, des coefficients ont été affectés à chaque nucléotide dégénéré selon qu'il substitue 1, 2 ou 3 nucléotides non dégénérés (Tableau 4). Le HyC étant le produit des coefficients de tous les nucléotides présents dans une séquence d'amorces, une formule a été définie :

$$HyC = 1^{ni} \times C_R^{nR} \times C_Y^{nY} \times C_M^{nM} \times C_K^{nK} \times C_W^{nW} \times C_S^{nS} \times C_B^{nB} \times C_H^{nH} \times C_D^{nD} \times C_V^{nV}$$

ni : nombre de nucléotides non dégénérés présent dans la séquence de l'amorce
 C_R^{nR} : Coefficient du nucléotide dégénéré R, nR : nombre de R dans la séquence

De cette formule se déduit alors,

$$HyC = 1^{ni} \times 2^{(nR+nY+nM+nK+nW+nS)} \times 3^{(nB+nD+nH+nV)}$$

Exemple de calcul du HyC d'une séquence dégénérée (Seq deg) :

Seq deg : 5' **C G C G G M T A C C G V A T Y G A G C** 3'

$$HyC = 1^{16} \times 2^{(0+1+1+0+0+0)} \times 3^{(0+0+1)}$$

$$= 1 \times 2^1 \times 2^1 \times 3^1$$

$$HyC (Seq deg) = 12$$

Les amorces non dégénérées ou dites simples ont alors un HyC = 1.

Les coefficients d'hybridation des amorces dégénérées ont ainsi été déterminés :

$$HyC (A7.d1) = 18 \quad HyC (A10.d1) = 141$$

$$HyC (Alb1-F) = 8 \quad HyC (Tlb1-R) = 32$$

Plus le HyC est élevé, plus l'amorce est dégénérée et donc moins stable et plus inefficace. Les tests PCR réalisés *in silico* ont montré que lorsque le HyC est supérieur à 200, cela affecte l'hybridabilité des amorces. Par conséquent, toutes les amorces dégénérées qui seront définies au cours de ce travail auront un HyC en dessous de 200.

Il est toutefois aussi possible qu'une très grande variation entre les différents HyC (ΔHyC) d'un même couple puisse avoir un effet sur l'efficacité des amorces lors d'une PCR. Le ΔHyC étant le différentiel entre les HyC de l'amorce 1 et l'amorce 2.

Les amorces ainsi définies ont été testées *in vitro* sur des souches modèles de lactobacilles.

1.2. Application sur des souches modèles du genre *Lactobacillus*

1.2.1. Tests PCR *in vitro*

Quatre souches du genre *Lactobacillus* décrites dans le § 1.2.1 du Chapitre Matériels et Méthodes ont été utilisées : la souche de référence *L. plantarum* WCSF1 c'est-à-dire celle dont l'ADN a servi pour la définition des amorces, la souche mère (*L. plantarum* NCIMB8826, souche ayant servi pour obtenir la colonie WCSF1, la souche de référence) ont été utilisées comme souches témoins positifs, la souche *Lactobacillus sakei-sakei* 16 ayant exprimée une activité biologique semblable aux témoins positifs (voir Chapitre 7) et les souches *Lactobacillus plantarum* ATCC 14917 et ATCC 8014 de la collection LMG qui ont été décrites comme ne possédant pas de gènes NRPS (Moleenar *et al.* 2005), comme témoins négatifs.

Les extractions des ADNs ont été réalisées comme décrites dans le § 2.1.1 du Chapitre Matériels et Méthodes. Les concentrations des ADNs ont été harmonisées entre 10-30 ng.μL⁻¹ pour la PCR.

Un test PCR sera dit positif lorsque, pour un couple d'amorce, la taille des fragments amplifiés sera similaire à celle obtenue lors des tests effectués *in silico*.

La figure 21 présente les résultats des PCR réalisés *in vitro* sur l'ADN des souches témoins positifs avec les amorces A7.1/T1.1, A10.4/C3.4 et A7.d1/A10.d1 et le tableau 8, l'ensemble des amplicons obtenus y compris avec les souches utilisées comme témoin négatif.

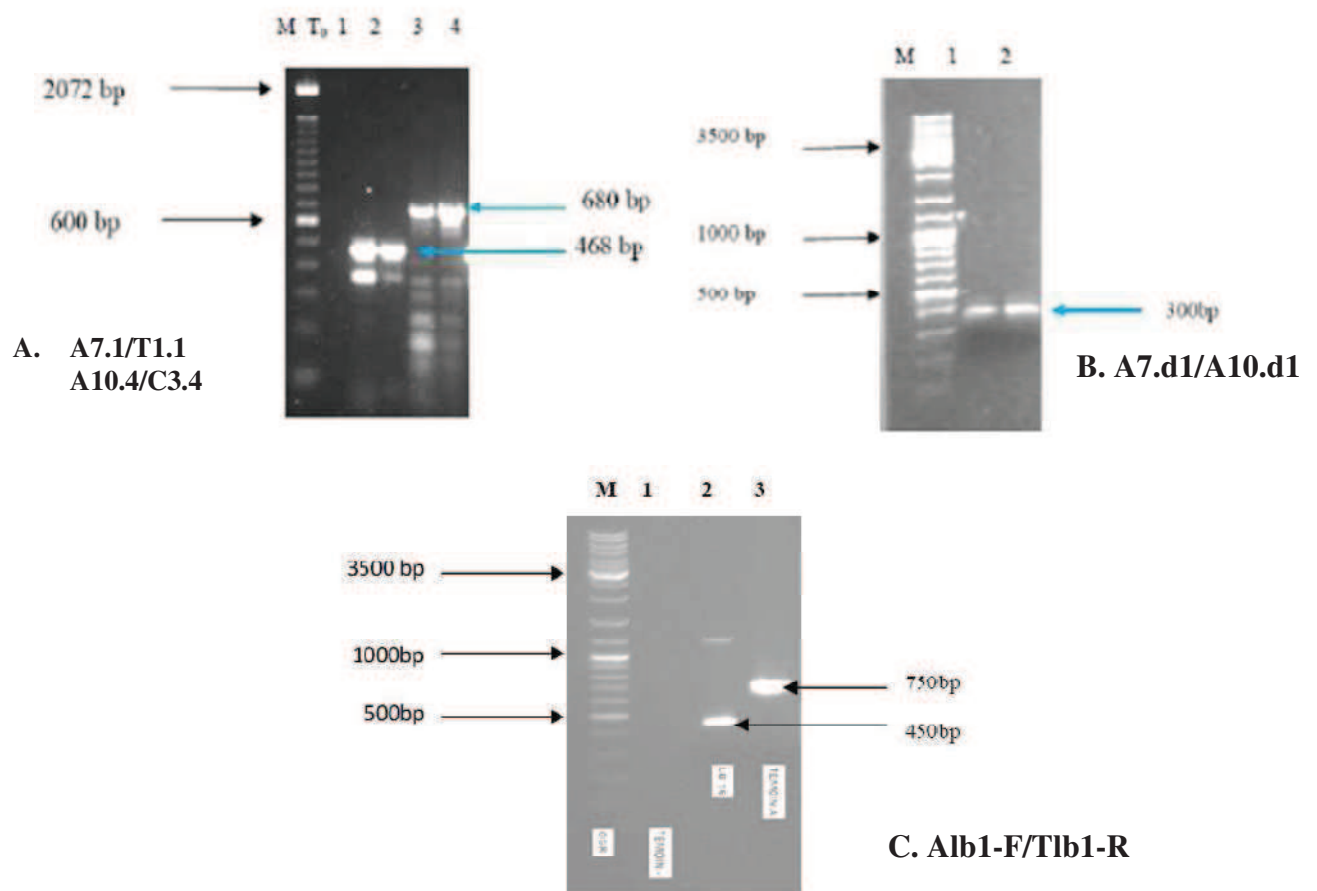


Figure 21. Détection sur gel d'agarose des fragments amplifiés par PCR sur les ADNs des souches témoin positif (**Gel A**: T0 témoin amorces sans ADN, Puits 1 : ADN *L. plantarum* WCSF1 avec amorce A7.1/T1.1, Puits 2 : ADN *L. plantarum* 8826 avec amorce A7.1/T1.1, Puits 3 : ADN *L. plantarum* WCSF1 avec amorce A10.4/C3.4, Puits 4 : ADN *L. plantarum* 8826 avec amorce A10.4/C3.4) (**Gel B** : Puits 1 : ADN *L. plantarum* WCSF1 avec amorce A7.d1/A10.d1, Puits 2 : ADN *L. plantarum* 8826 avec amorce A7.d1/A10.d1) (**Gel C**. Puits 1 : témoin sans ADN, Puits 2 : DNA *L. sakei-sakei* 16 avec les amorces Alb1-F/Tlb1-R, Puits 3 : DNA WCSF1 avec les amorces Alb1-F/Tlb1-R) M : Marqueur de Taille O' Gène Ruler

Tableau 8. Taille des fragments obtenus par PCR avec les 4 couples d'amorces définies sur l'ADN de 5 souches de lactobacilles

Souches de lactobacilles	Amorces testées	Tailles obtenues (bp)
<i>Lactobacillus plantarum</i> WCSF1	A7.1/T1.1	350, 468
	A10.4/C3.4	680, *trainée
	A7.d1/A10.d1	300
	Alb1-F/Tlb1-R	750
<i>Lactobacillus plantarum</i> 8826	A7.1/T1.1	350, 468
	A10.4/C3.4	680, *trainée
	A7.d1/A10.d1	300
	Alb1-F/Tlb1-R	750
<i>Lactobacillus sakei-sakei</i> 16	A7.1/T1.1	380, *trainée
	A10.4/C3.4	Pas de bandes
	A7.d1/A10.d1	300, 400
	Alb1-F/Tlb1-R	450
<i>Lactobacillus plantarum</i> ATCC 14917	A7.1/T1.1	Pas de bandes
	A10.4/C3.4	350, 450, 800, 1400
	A7.d1/A10.d1	800, 1200
	Alb1-F/Tlb1-R	> 1500
<i>Lactobacillus plantarum</i> ATCC 8014	A7.1/T1.1	Pas de bandes
	A10.4/C3.4	180, 350, 450, 800, 1400
	A7.d1/A10.d1	800, 1200
	Alb1-F/Tlb1-R	> 1500

*trainée = étalement non séparable de bandes, > = supérieur à

Le couple d'amorce A7.1/T1.1 a amplifié des fragments de tailles attendues (468 bp) dans les deux souches témoins positifs *L. plantarum* WCSF1 et NCIMB 8826. Ces mêmes amorces ont aussi amplifié des fragments de tailles non attendues de 350 bp respectivement chez les souches témoins positifs et de 380 bp chez la souche *Lactobacillus sakei-sakei* 16. Aucun fragment n'a été détecté par ces amorces chez les souches témoins négatifs *L. plantarum* ATCC 14917 et 8014 comme attendu.

Le couple A10.4/C3.4 a détecté des fragments de tailles attendues (680 bp) dans les deux souches témoins positifs et, en plus une trainée de bandes. Aucun fragment n'a été amplifié par ces mêmes amorces chez *Lactobacillus sakei-sakei* 16. Elles ont, en outre, détecté des amplicons de tailles variant de 180 bp à 1500 bp chez les souches témoins négatifs *L. plantarum* ATCC 14917 et 8014 alors qu'on n'attendait aucune détection de bandes.

On peut affirmer que les amorces non dégénérées A7.1/T1.1 et A10.4/C3.4 ont été non-sélectives car elles ont amplifié des bandes non attendues chez la souche de référence *L. plantarum* WCSF1 (ayant servi à définir ces amorces).

Le couple d'amorce dégénérée A7.d1/A10.d1 a amplifié des fragments de tailles attendues (300 bp) respectivement dans les deux souches témoins positifs et chez la souche *Lactobacillus sakei-sakei* 16. Elles ont, en outre, détecté un autre fragment non attendu de 400 bp chez *Lactobacillus*

sakei-sakei 16. Des fragments non attendus ont été aussi amplifiés par les amorces A7.d1/A10.d1 chez les souches témoins négatifs.

La bande de 750 bp a été amplifiée par les amorces Alb1-F/Tlb1-R comme prévue dans la souche témoin positif *L. plantarum* WCSF1. Ces amorces ont aussi détecté un fragment d'environ 450 bp différent de la taille attendue avec l'ADN de la souche *L. sakei-sakei* 16.

Au regard de ces résultats, les amorces dégénérées A7.d1/A10.d1 et Alb1-F/Tlb1-R sont spécifiques lorsqu'elles sont appliquées sur les souches témoins positifs alors que celles non dégénérées, par l'amplication de fragments non attendus sur les souches témoins positifs apparaissent moins spécifiques.

1.2.2. Séquençage de fragments amplifiés par PCR

Des fragments amplifiés par les amorces dégénérées ont été clonés puis séquencés comme décrit dans les § 2.1.6 à 2.1.9 du Chapitre Matériels et Méthodes.

Les amplicons de 750 bp et 450 bp détectés par les amorces Alb1-F/Tlb1-R respectivement chez *L. plantarum* WCSF1 et *L. sakei-sakei* 16, puis le fragment de 300 bp mis en évidence par les amorces A7.d1/A10.d1 chez *L. sakei-sakei* 16 ont été séquencés.

Le choix de séquencer le fragment de 750 bp permettra de confirmer l'amplification de gènes NRPS de WCSF1 par les amorces Alb1-F/Tlb1-R. L'amplification des amplicons de 450 bp (fragment de taille non attendue avec les amorces Alb1-F/Tlb1-R) et de 300 bp (fragment de taille attendue avec les amorces Alb1-F/Tlb1-R) permettra de savoir si ces amplicons sont issus de gènes NRPS.

La figure 22 présente les séquences nucléiques des trois fragments décrits ci-dessus.

A.

Tlb1-R

TGACCCGCTAGCGACATCCTTCCCTGCCAATTGAAAAGATTGAAGATAGTGTAGGACTATAT
ATTCGAACGCTACCAATTCAATTTAGTGCACACTGATAAGTCGATTAACTATCAATTGCATAA
ATTGCAGGATATTAATAATGAGTGCATGTGGCTGGCTTACAGAGAAATGGAACGCGATTATT
TGACACATTGTTTGTGTTATGAAAACCTATCCGGTATCTACTGATAGTTTCGAGCAGTATGCTAA
AGGTGCACAATTTTCGTGCACAAGAAAACTAGATTATCCCTTAACTGTTATGTTGTCAGAG
TTATCAGATAGTTTAAGTTTTAGTTTAATGTTTGGCTGGTGAAATTTTTGCGGAAGCGACAAT
TGATCACCTATTTGCGATGGTTCAGCGGTTAGTATCACAGATTGTAAAGGGAGTCAATAGCT
TTACCTTTTATTGAAAAGGCTCCAACTTTGGTACTGCTTATTATCCTGAGAAGACTATCGCT
GAAGAATTTGAGCAGCAGGTCCGTGAACATCCTAATAATATTGCCCTTGATTTCAAGAATAT
CCGTATACCTTTGAAGAGTTAAATCAAAGAGCAAATAGGGTGGCACATACGCTGGTTGAAG
ATTATGATATTAAGCTAGGTGATCGAGTACCGTTACTGCTGCCTAAAAGTGAGAATACCATT
ATTGCCATTTTGGCGATTTTGAAGCGGGCGCAGTTTACGTTCCGATGGCAGGCTAAGATTG
CTCGAACAGCAAATAAGCTACTGGAATAATAGTGCCAACAGAGA

Alb1-F

B.

Tlb1-R

TGATCTGGTAGCGACATCCTTCCCTGTAATTATGGTTAAATATCATTATACCCAACCTAGATC
CGTTAGAAATGAAATGATACCTTTATTAATACCGGTAACAGTCTAAAAAGTCCAGAACGCG
CAATTTTCATCAACGTTCTGGATCTTCAACCGAATTATTTTCGTTTGCTTCTGCTTACTGAATC
GTTGTCGTCCCCAAACAAACAGCAGGACATAAATAACGATAATCACCAAATTATAAACTAAA
CTAATTATCAGTTTCCGAACGCCTTTACCCGGTACCAGTACAAAGACTAATCAGACAATAAA
GTAACCAACTGGTATAACACTACCGACTATGACAATATTAGTCCATGTGTCGTAGTTTGTGT
CATCACAGCTAAAAATTGCCCGAACGACAAAGTAAGCTATTGGAAGAATAGTGCCAACAGAAA

Alb1-F

C.

A7.d1

GTAGAATTTATRVDACNGGDGATCGTCGATCAGCTTACCATTTAACTTCCGTCGGCATTCTT
AAAACCTAATTTTTTTCATCTCCTTTGGCTTCCTTTTCATATAGATGTTTAAATTCAATGTCA
TGACCATCTAGCCGGCCTTCTACCTTAATGACCCGATTTTCAATCGCATTCATTGCTTCAGC
ATTTTGCTGCCTCACTTTTAACTTTTCATCGGTAAAATGGCTCAGCCGCTTGCTTAAATCGT
TAAGCGGAATACGAACCGTTTTATTAAATCCAATTAGCCAATACACAAATACTAGTGACA
ATGGCAAACCTCGATCGACGTGATCVADYTTDCCATT

A10.d1

Figure 22. Séquences des fragments amplifiés par les amorces dégénérées (A. Fragment 750 bp de *L. plantarum* WCSF1 avec les amorces Alb1-F/Tlb1-R, B. Fragment 450 bp de *L. sakei-sakei* 16 avec les amorces Alb1-F/Tlb1-R, C. Fragment 300 bp amplifié par les amorces A7.d1/A10.d1 chez *L. sakei-sakei* 16)

Les fragments ainsi séquencés ont été analysés par Blast N afin de déterminer les homologies avec les gènes présents dans la banque Genbank de NCBI.

Le fragment 750 bp (figure 22 A) amplifié par les amorces Alb1-F/Tlb1-R a une taille de 789 bp. Il est identique à 100% avec une E. value = 0.0 à un fragment du gène *npsA* de *Lactobacillus*

plantarum WCSF1. On peut conclure sur l'efficacité de ces amorces à reconnaître le gène *npsA* comme attendu.

Le fragment de 450 bp (figure 22 B) séquencé mesure en réalité 435 bp, une taille différente de celle attendue qui oscille autour de 750 bp. L'analyse Blast N montre qu'il a un pourcentage de similarité de 85% avec une E. value = $8e-101$ à un gène codant pour une protéine membranaire de *L. plantarum* WCSF1. Ce gène est non NRPS. La détection de ce gène corrobore la taille de l'amplicon détecté non attendu pour ces amorces. En effet, l'amplification de fragments de taille non attendue augure de la nature non NRPS du gène détecté. Une des hypothèses de l'amplification de gènes non NRPS pourrait s'expliquer par l'absence de gène NRPS dans cette souche et de la nature dégénérée de ces amorces. En effet lors d'une PCR, ce type d'amorces a tendance à s'hybrider à d'autres gènes lorsque ceux recherchés sont absents du génome.

L'amplicon de 300 bp (figure 22 C) amplifié par les amorces A7.d1/A10.d1 a une taille réelle de 338 bp. L'analyse par Blast N montre une similarité soit de 92% avec la protéine 59 du prophage Lp1 ou de sa protéine 60, soit à 84% avec la protéine 57 du prophage Lp2 de *Lactobacillus plantarum* WCSF1.

Pour déterminer si les protéines 59 ou 60 de prophage Lp1 obtenues avec une forte similarité (92%) sont de type non ribosomal, nous avons comparé leurs séquences protéiques respectives aux domaines d'adénylation protéique (zone d'amorçage des amorces A7.d1/A10.d1) de la synthétase NpsA. A l'aide du logiciel Vector NTI, nous avons réalisé un alignement local entre ces protéines, puis à partir de leur position consensuelle, des pourcentages de similarité ont été déterminés en considérant uniquement les séquences qui se sont appariées (Tableau 9).

Tableau 9. Pourcentage de similarité entre les domaines d'adénylation de NpsA et des protéines de prophage détectées par les amorces A7.d1/A10.d1 (Vector NTI)

Séquences protéiques	A1 NpsA	A2 NpsA	A3 NpsA	A4 NpsA	A5 NpsA
Protéine 59 prophage Lp1	26,7%	23,5%	19,4%	17,9%	20,8%
Protéine 60 prophage Lp1	21,0%	26,1%	26,1%	21,4%	16,8

Cette analyse montre que le fragment amplifié de la souche *Lactobacillus sakei-sakei* 16 a un pourcentage de similarité entre 26,66% et 27,44% comparativement aux domaines d'adénylation de la synthétase NpsA. On peut affirmer que ce fragment de taille attendue détecté par les amorces A7.d1/A10.d1 n'est pas de type NRPS au vu de cette faible similarité.

Conclusion sur la définition des amorces

On peut conclure que les amorces non-dégénérées A7.1/T1.1 et A10.4/C3.4 n'ont pas permis de détecter uniquement le gène *npsA* chez la souche de référence comme attendu car elles ont conduit à une formation de trainées. Le couple dégénéré A7.d1/A10.d1, bien qu'ayant détecté uniquement le gène *npsA* chez *Lactobacillus plantarum* WCSF1, a mis en évidence d'autres gènes non NRPS (cas de *Lactobacillus sakei-sakei*). Même si cette amplification au niveau de la souche *Lactobacillus sakei-sakei* 16 peut s'expliquer par le fait que cette souche ne possède probablement pas de gènes NRPS, cela reste néanmoins confus quant à la sélectivité de ces amorces.

On peut alors dire que les amorces dégénérées Alb1-F/Tlb1-R qui ont été définies à partir de l'alignement des séquences nucléiques des domaines d'adénylation et de thiolation par rapport aux amorces A7.d1/A10.d1 (alignement des séquences nucléiques des modules de NpsA et NpsB) ont été plus efficaces. En effet, à l'opposé du couple A7.d1/A10.d1, les amorces Alb1/Tbl1 ont détecté un amplicon de taille non attendue confirmant que celui-ci serait non NRPS après le séquençage.

On peut alors affirmer que les amorces Alb1-F/Tlb1-R sont plus sélectives que les amorces A7.d1/A10.d1, sachant que les deux couples ont mis en évidence le gène *npsA* dans la souche de référence comme prévu.

La stratégie qui consiste à définir des amorces dégénérées à partir des séquences nucléiques des domaines d'adénylation et de thiolation serait donc la plus intéressante.

II. Criblage de bactéries lactiques pour la recherche de molécules actives non ribosomiales

Les amorces Alb1-F/Tlb1-R ont alors été utilisées pour rechercher des gènes NRPS dans une collection de bactéries lactiques, au préalable, les activités biologiques de ces souches ont cependant été déterminées afin de sélectionner les souches présentant un potentiel de synthèse de peptides actifs pouvant être d'origine non ribosomiale.

2.1. Recherche d'activités biologiques

Une collection de bactéries lactiques composée de plusieurs souches isolées de *Clupea harengus* fumé (collection ProbioGEM) a été criblée. On y retrouve les espèces *Lactobacillus plantarum*, *L. casei*, *L. sakei-sakei*, *L. pentosis*, *L. sakei-curvatus*, *L. pentosus* et *Lactococcus lactis*. A chaque souche a été affecté un code spécifique.

Ces bactéries ont été criblées pour déterminer leurs éventuelles activités antifongiques.

Les différentes souches ont été testées pour leur capacité à inhiber la croissance de spores de moisissures.

Différentes suspensions de spores concentrées à 10^4 cellules/mL des souches de *Penicillium roquefortii*, *Mucor rouxii*, *Rhizopus nigricans* et *Fusarium oxysporum* ont été réalisées comme décrit dans le § 2. 3. 3 du Chapitre Matériels et Méthodes.

Ces suspensions de spores ont alors été utilisées pour cribler les souches issues de la collection de poissons fumés et des souches témoins *Lactobacillus sakei-sakei* 16, *Lactobacillus plantarum* WCSF1 et NZ 1400. En effet, des travaux antérieurs (non publiés) au sein de notre laboratoire avaient montré que *Lactobacillus sakei-sakei* 16 et *Lactobacillus plantarum* WCSF1 ont des activités antifongiques similaires vis-à-vis d'un certain nombre de moisissures.

Le tableau 10 présente les activités antifongiques de bactéries lactiques isolées du hareng fumé.

Tableau 10. Activités antifongiques de bactéries lactiques isolées de *Clupea harengus* fumé

Souche	<i>R. nigricans</i>	<i>P. roquefortii</i>	<i>M. rouxii</i>	<i>F. oxysporum</i>
Lb 16	+++	++	+	+++
LbWCSF1	+	-	+	+++
Lb NZ1400	+++	-	+	-
Lc A12	+	+++	-	++
Lb 51	+	+	+	+
Lb 53	++	++	+	++
Lb 54	-	-	++	+
Lb 57	+	+	+	++
Lb 62	++	+	++	++
Lb 63	+	+	+	++
Lb 65	-	+	-	-
Lb 67	+	+	+	+
Lb 68	+	+	+	++
Lb 70	++	+	+	+
Lb 71	+	+	+	++
Lb 94	-	-	-	-
Lb 95	-	-	-	-
Lb 96	+++	+++	+	+++
Lb 97	+++	+++	+	+++
Lb 98	+++	+++	++	+++
Lb 99	-	-	-	+++
Lb 100	-	-	-	+++
Lb 101	-	-	-	-
Lb 102	+++	+++	++	+++
Lb 104	-	-	+	+++
Lb 106	-	-	-	-
Lb 108	-	-	-	-
Lb 110	+++	+++	+	+++
Lb 111	+++	+++	-	+++
Lb 112	-	-	+++	++
Lb 113	-	-	-	+++

-: pas d'inhibition, + : zone d'inhibition entre 0,1 et 0,5 cm, ++ : zone d'inhibition entre 0,5 et 1cm, +++ zone d'inhibition supérieure à 1cm, (Lb : *Lactobacillus plantarum*, Lc : *Lactococcus*).

L'inhibition par certaines souches de la croissance de certaines moisissures pourrait être due à la synthèse de métabolites secondaires.

La recherche des activités antifongiques dans cette collection montre que ces activités peuvent différer fortement d'une souche à l'autre.

2.2. Recherches de gènes NRPS dans la collection de bactéries lactiques

Les amorces Alb1-F/Tlb1-R définies au niveau du gène *npsA* de *L. plantarum* WCSF1 ont été utilisées pour le criblage de la collection issue de poissons.

La souche *Lactobacillus plantarum* WCSF1 a été utilisée comme témoin positif. La figure 23 montre quelques fragments obtenus lors du criblage de ces souches par PCR.

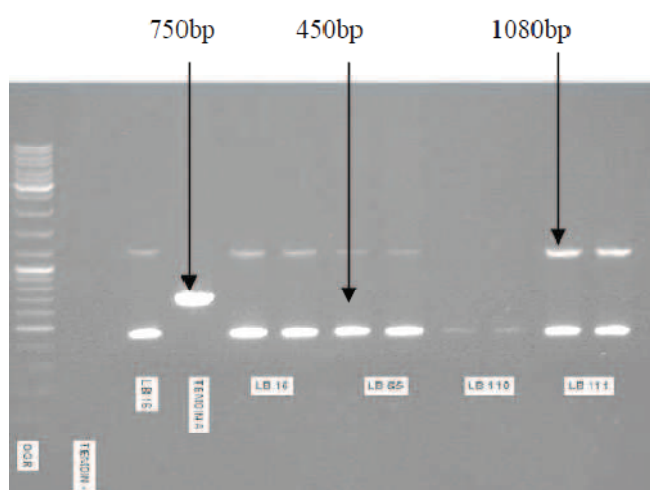


Figure 23. Détection sur gel d'agarose des fragments amplifiés par PCR effectués sur les ADNs des souches 16, 65, 110 et 111 avec les amorces Alb1-F/Tlb1-R (Témoin: *L. plantarum* WCSF1)

Chez la plupart des lactobacilles issus de poissons fumés, ces amorces ont conduit à un profil presque similaire c'est-à-dire à l'amplification de deux fragments non attendus de 450 bp et 1080 bp. Ceci suggère l'absence de gènes NRPS chez les souches étudiées. On notera qu'aucun fragment n'a été détecté chez l'espèce *L. lactis* A12.

Afin de vérifier la qualité non NRPS des fragments amplifiés, un d'entre eux a été séquencé (fragment de 1080 bp de Lb 111, figure 24).

Tlb1-R
TGACCTGCTAGCGACATCCTTGC GAATTGCAATTAGCCGGGACACTGATGCCGATGACCAGC
GGGTAATAACCTTCAAGCCCATTTGGTGAACACTATCAACGACTGGTTGATCAATTGAAGGAG
TGACAACCATGGCAGAAGAAGTTGTTGATGTACGGCCTGCTGAAGTCGCGGATGCAGCACAA
TTATTAGCGCTACTAGCTCAGCTTGGACGTGAAAGTAATACTTTCACTGTTGATGAAGGCAT
TGAGGATTTGAGTGAAACGGATGAACAGGCGCAAATTGAGCGIATCAATGGCACAACGACCA
ACATTATTTTCGTTGCGACATTAGATTACGGTTGATTGGTGTGAGCACCGTGCAAGCTAGT
ACGGATTTTAGTGCGGCGCAAGGTGAAGTTGGCGTAGCGGTACTTAAAGAAATTTTGGGGCAT
GGGTCTGGGGACTGCTTTGATTGAGGAAGTCTTAGATTGGGCGCGTAACTATAGCTCATTGG
AACGCCCTCGTGTGACGGTCCAGTTACGAAATGTGCGTGCCGTGAAGCTGTACCAGCAICTT
GGGTTCGAAAATTGTACGGCAACGAGCTATGATGTGCTTGACCCAACCGGCAAACGGGTTCC
AGCTATTTGATATGGGAATCTGGGTTTGAAATCCGGGATTCGCTGAACTGACGGCAAACAGGT
TCTGATAGTAGATGGGCCGGAATTTAAGCGTTGACGAACCAGGTCTAACCGGCGACGGCTTG
TGAATGATTGGCTCATCTCCAGTCGGTTGGGGGAATTGATAAGCTGTCTACGAGTATTGAT
AATCGGTAGCGATGCGTAAACAAAAAGTCCAGCATCTTATTGCTGGACTTGCATAAAATTTA
ATTTTAAACCCCTGAACGCGAACGAAAGGTTTGAGTTGCTCAGTTCCTAAAAGCTTGCCTTCGG
TGAGTAAAATGTTGGCGCAAGCTAGGCTGTCTAGTGCATTATGATGATGCTGCAGGTG
ATATCCAGTGCTTGACAGACTGTATCCAGCTTGTGATTAGGCAATTCCTGGATAGAACCCTTG
ACTGGTCTTGACGGTATCCATGGATAGGTACTCCGGCGCGCTGATGCCGTAGTGGGCAAGCG
TATTGCGTAAACACACCAATATCAAAAGGCGCGTTGTGCGCAATTACGAGGCGGTTCGCGTTGG
AATAATAGTGCCAACAGAAA

Alb1-F

Figure 24. Séquences du fragment de 1080 bp amplifié par les amorces dégénérées Alb1-F/Tlb1-R de la souche Lb111

Ce fragment mesure 1180 bp et est similaire à 98% avec une E. value = 0.0 à celui responsable de l'ATPase ou d'une kinase (putative) acétyltransferase (putative) de *L. plantarum* JDM1 ou de WCSF1.

L'analyse comparative de la séquence protéique de cette acétyltransferase avec les différentes zones d'hybridation dans les domaines A et T montre que la similarité oscille entre 19% et 25%. Cette faible similarité indique que ce fragment n'est donc pas d'origine non ribosomiale et confirme que lorsque la taille du fragment amplifié est non-attendue, la portion de gène amplifié est non NRPS. Les amorces dégénérées conduiraient donc à une amplification d'autres types de fragments lorsque celui recherché est absent.

Conclusion partielle sur le criblage des bactéries lactiques

Au terme de ce criblage moléculaire, on remarque que les amorces Alb1-F/Tlb1-R (définies à partir du gène *npsA* de *L. plantarum* WCSF1) n'ont détecté aucun fragment de type non ribosomal dans toutes les souches testées. Cela peut être interprété par l'absence du gène *npsA* voire même très probablement de gènes NRPS dans les souches de bactéries lactiques excepté chez l'espèce WCSF1, confirmant ainsi les travaux de Molenaar *et al.* (2005).

L'absence des gènes *npsA* et *npsB* dans toutes les souches de *Lactobacillus plantarum* étudiées à ce jour et la non-détermination de la fonction de la synthétase codée par ces gènes sont probablement le fait que ces gènes aient été acquis par l'espèce WCSF1 lors d'un phénomène de transfert horizontal (THG). En effet, Daubin *et al.* (2003) ont montré que 70% des protéines codées par des gènes issus d'un transfert horizontal ont une fonction inconnue. De plus tout récemment, Zhang *et al.* (2009) en séquençant *Lactobacillus plantarum* JDM1 (proche à plus de 90% au niveau nucléique de *L. plantarum* WCSF1), ont montré que celui-ci ne contient aucun gène codant pour des synthétases non ribosomiales c'est-à-dire ni les gènes *npsA* et *npsB* ni aucun autre gène de ce type.

Une analyse bioinformatique a alors été réalisée à l'aide de l'outil GEECEE Mobyly (Institut Pasteur Paris, France) pour vérifier les proportions des nucléotides G et C des gènes *npsA* et *npsB* par rapport au génome de *L. plantarum* WCSF1. Le pourcentage en GC des gènes *npsA* et *npsB* est respectivement de 36% et 37% alors que celui du génome de *L. plantarum* WCSF1 est de 44%. Il est très fort probable au vu de cet écart de proportion, que la présence des gènes *npsA* et *npsB* chez *L. plantarum* WCSF1 pourrait être le résultat d'un transfert horizontal de gènes.

Ces résultats tendent à montrer que les activités biologiques de toutes les souches analysées excepté *Lactobacillus plantarum* WCSF1 ne sont pas dues à des molécules NRPS. On peut cependant affirmer que les activités antifongiques de WCSF1 notamment anti- (*M. rouxii*, *R. nigricans* et *A. niger*) ne sont pas non plus le fait d'une molécule qui serait induite par la présence du gène *npsA* car la souche *L. plantarum* NZ 1400 (mutant *npsA*⁻ de WCSF1) possède les mêmes propriétés. Il serait cependant hasardeux d'étendre ces résultats à l'ensemble des bactéries lactiques. En effet, la présence éventuelle de plusieurs enzymes impliquées dans la biosynthèse de molécules non ribosomiales ou de polykétides synthases a été récemment montrée chez certaines bactéries lactiques comme *Lactococcus lactis* str. KF282 (Siezen *et al.* 2008). Bien que préliminaire, l'identification de la molécule correspondante et l'étude de leur fonction par analyse comparative avec certains ORFs de molécules de bacilles fait conclure de façon prédictive qu'il pourrait s'agir soit d'un sidérophore catécholate de type entérobactine (Tor *et al.* 1992) ou soit de la bacillibactine (Dertz *et al.* 2006) avec une similarité variant de 33-53%.

Conclusion et perspectives

Le criblage moléculaire d'une collection de bactéries lactiques présentant des activités biologiques n'a pas permis de mettre en évidence des souches productrices de molécules non ribosomiales. La différence de sélectivité des amorces ont démontré la difficulté de définition de ces dernières pour obtenir, par exemple, des amorces universelles qui serait alors capables de détecter tous les gènes de types NRPS dans tous les genres bactériens.

Une alternative pourrait être de définir plusieurs couples d'amorces en tenant compte de l'espèce et du genre bactérien, des synthétases de type NRPS (séquences conservées) impliquées dans la synthèse de molécules connues, de l'homologie entre les opérons notamment au niveau nucléique mais aussi du niveau de dégénérescence de ces oligonucléotides.

Chapitre 6 Détection par PCR des gènes non ribosomiaux codant pour la biosynthèse de lipopeptides de *Bacillus* sp.

La définition d'amorces universelles pour la détection de gènes NRPS apparaît peu réaliste comme le montre le chapitre précédent. Une nouvelle stratégie orientée vers la définition d'amorces spécifiques de certains types de molécules a alors été mise au point.

Dans ce chapitre, nous présentons dans un premier temps, la mise au point de cette stratégie qui consiste à déterminer des amorces à partir de séquences nucléiques des domaines d'adénylation et de thiolation des synthétases de molécules connues et produites par des souches du genre *Bacillus*, puis en deuxième partie, une procédure de validation de cette stratégie par l'application des amorces ainsi définies sur des souches de *Bacillus* connues ou non pour leur production de molécules non ribosomiales.

I. Stratégie de la mise au point des amorces spécifiques

Dans le but de définir des amorces spécifiques, les opérons des 3 familles de lipopeptides NRPS du genre *Bacillus* (voir Chapitre 2) ont été utilisés. Afin d'étudier les homologies entre les ORFs qui codent pour les synthétases de lipopeptides de la même famille, une comparaison a été réalisée au niveau protéique et nucléique.

1.1. Comparaison des opérons

Ainsi, les ORFs de l'opéron surfactine ont été comparés à ceux de la lichenysine, ceux de la plipastatine à la fengycine et enfin les ORFs de mycosubtiline contre ceux de l'iturine A à l'aide de Needle EMBOSS de EMBL-EBI (European Bioinformatics Institute).

Les résultats des comparaisons effectuées sont consignés dans le tableau 11.

Tableau 11. Analyse comparative des synthétases de lipopeptides NRPS de souches de *Bacillus* par Needle

Origines des séquences	Gènes de lipopeptides	Similarité protéique (%)	Similarité nucléique (%)
Synthétases de la surfactine de <i>B. subtilis</i> 168 et de la lichenysine de <i>B. licheniformis</i> ATCC14580.	<i>srfA-A/lch A</i>	77.2	64.7
	<i>srfA-B/lch B</i>	77.1	65.5
	<i>srfA-C/lch C</i>	75.1	62.9
	<i>srfA-D/lch D</i>	75.6	62.8
Synthétases de la fengycine de <i>B. subtilis</i> F29-3 et de la plipastatine de <i>B. subtilis</i> 168	<i>ppsA/fen C</i>	87.6	43.4
	<i>ppsB/fen D</i>	89.6	44.3
	<i>ppsC/fen E</i>	90.0	43.1
	<i>ppsD/fen A</i>	88.1	42.9
	<i>ppsE/fen B</i>	88.4	42.8
Synthétases de la mycosubtiline de <i>B. subtilis</i> 6633 et de l'iturine A de <i>B. subtilis</i> RB14	<i>mycA/ituA</i>	88.7	80.7
	<i>mycB/ituB</i>	88.6	79.8
	<i>mycC/ituC</i>	79.0	69.7

Protéines (Method: Emboss/ Matrix: Blosom62/Gap Open: 10, Gap Extend 0.5) DNA (Method: Emboss/ Matrix: DNA full/Gap Open: 10, Gap Extend 0.5).

Au niveau protéique, les ORFs des différents opérons sont globalement similaires entre eux à l'intérieur de chaque famille. Ils ont un pourcentage de similarité d'environ 76%, 88% et 86% respectivement au niveau des familles de surfactines, de fengycines et d'iturines. Ces similarités élevées entre les ORFs confirment que ces enzymes synthétisent des molécules de structures similaires.

Au niveau nucléique, les synthétases de la famille des fengycines se distinguent des deux autres familles par une similarité plus faible environ 43%, contre respectivement en moyenne 63% et 77% pour les ORFs des surfactines et des iturines. Le pourcentage de similarité moyen entre les ORFs des iturines masque la disparité entre ces ORFs. En effet, les ORFs *mycC* et *ituC* ont une relative faible homologie (69,7%) contrairement aux autres ORFs de la même famille (environ 88%). Cela peut s'expliquer par une inversion des modules 6 et 7 qui induit à une inversion des acides aminés dans les peptides produits (Ser-Asn pour la mycosubtiline et Asn-Ser pour l'iturineA) et démontrée par Tsuge *et al.* (2001) entre les domaines. Une comparaison au niveau protéique des domaines d'adénylation activant à la fois les mêmes acides aminés (Ser ou Asn) ont montré qu'ils ont une très forte homologie (>80%).

Les niveaux de similarité entre les ORFs des synthétases d'une même famille a conduit à la définition de quatre couples d'amorces susceptibles de détecter la présence des 3 familles dans le genre bactérien. En effet, la faible similarité observée au niveau des séquences nucléiques des ORFs de la plipastatine et de la fengycine (environ 43%) ne permet pas de définir un unique

couple d'amorce capable de détecter ces deux gènes dans des souches bactériennes contrairement aux autres gènes issus des surfactines et des iturines.

1.2. Définition des amorces

Le principe est de définir des amorces spécifiques à partir des opérons séquencés de lipopeptides NRPS produits par l'espèce *Bacillus subtilis* décrits dans le Chapitre 2. Le tableau 12 présente le potentiel de synthèse de lipopeptides connus à ce jour chez les souches de *Bacillus subtilis* et les opérons non ribosomiaux correspondants qui ont été déterminés.

Tableau 12. Potentiel de synthèse de lipopeptides non ribosomiaux et présence de gènes correspondants mis en évidence chez les souches de *Bacillus subtilis*

	<i>B. subtilis</i> 168	<i>B. subtilis</i> 21332	<i>B. subtilis</i> 6633	<i>B. subtilis</i> 9943	<i>B. subtilis</i> F29-3
Molécules synthétisées	aucune	Surfactines Plipastatine	Surfactines Mycosubtilines	Surfactines Fengycines Iturines	Fengycine
Présence de gènes ou opérons	Surfactine Plipastatine	Non déterminé	Mycosubtiline	Non déterminé	Fengycine

Des amorces ont été définies à partir des opérons surfactine et plipastatine de *Bacillus subtilis* 168, de l'opéron fengycine de *Bacillus subtilis* F29-3 et de l'opéron mycosubtiline de *Bacillus subtilis* ATCC 6633 selon la démarche décrite dans le § 2.2.2 du Chapitre Matériels et Méthodes. Ces lipopeptides représentent les trois familles de lipopeptides NRPS produites par le genre *Bacillus*.

Les alignements de séquences nucléiques des domaines d'adénylation et de thiolation des ORFs des synthétases SrfA-A et SrfA-B ont été effectués pour la définition des amorces As1-F/Ts2-R, des ORFs des synthétases MycA, MycB et MycC pour le couple Am1-F/Tm1-R, des ORFs des synthétases PpsA, PpsB, PpsC, PpsD et FenB, FenD, FenE, FenA respectivement pour la définition de amorces Ap1-F/Tp1-R et Af2-F/Tf1-R.

Les ORFs des synthétases SrfA-C, de PpsE et FenC ont été exclues de ces alignements car la présence de toutes les ORFs ne permet pas d'obtenir des motifs conservés susceptibles d'être utilisés pour définir des amorces spécifiques. Le choix d'éluder les domaines d'adénylation et de thiolation issus de ces synthétases s'est fait sur le critère que leur délétion permet d'avoir des motifs conservés selon le logiciel Vector NTI. En effet, le but des alignements est d'obtenir des motifs conservés pouvant servir pour la définition des amorces.

La figure 25 présente les alignements réalisés et les motifs qui ont servi à la définition des couples d'amorces de la mycosubtiline.

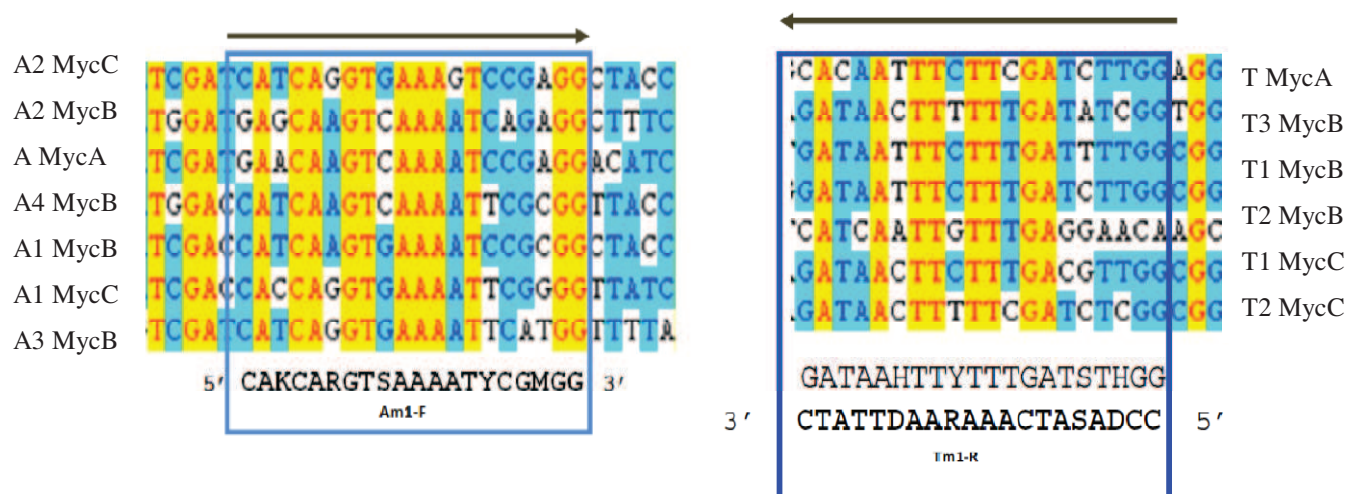


Figure 25. Représentation schématique des alignements de séquences nucléiques et des motifs conservés (encadrés en bleu) qui ont servi à la définition des amorces spécifiques Am1-F et Tm1-R

A l'aide du logiciel Vector NTI, un degré de consensus ou 'consensus position' est déterminé sur la sélection de la zone souhaitée. Ce degré de consensus permet de jauger de la conservation du motif choisi pour définir l'amorce. Il est exprimé en pourcentage. Le motif dit 'idéal' sera alors celui qui aura un degré de consensus le plus élevé comme ceux présentés dans la figure 25 précédente. Les autres couples d'amorces ont aussi été définis de la même manière.

1.3. Calcul du Coefficient d'hybridation (HyC)

Le coefficient d'hybridation (HyC) de toutes les amorces a été calculé comme décrit dans le § 1.1.2 du Chapitre 5. Le tableau 13 montre les caractéristiques des amorces qui ont été déterminées : la position sur les différents ORFs, la taille des séquences des amorces et les numéros d'accèsion Genbank des synthétases utilisées puis leurs origines.

Tableau 13. Caractéristiques des amorces spécifiques pour la détection des synthétases de lipopeptides de *Bacillus*

Lipopeptides	Abréviation	Séquences 5' 3'	Taille (bp)	HyC	*Positionnement des amorces sur les séquences alignées	Origine bactérienne des gènes	Numéro accession Genbank des synthétases
Fengycines	Af2-F	GAATAYMTCGGMCGTMTKGA	20	34	1138 -1157	<i>B. subtilis</i> F29-3	AF023464
	Tf1-R	GCTTTWADKGAATSBCCGCC	20	72	75-94		BACFENB CAA09819 AF023465
Plipastatines	Ap1-F	AGMCAGCKSGCMASATCMCC	20	64	693-713	<i>B. subtilis</i> 168	CAB13717.1
	Tp1-R	GCKATWWTGAARRCCGGCGG	20	32	69-89		CAB13716.1 CAB13715.1 CAB13714.1
Surfactines	As1-F	CGCGGMTACCGVATYGAGC	19	12	1182-1200	<i>B. subtilis</i> 168	D13262.1
	Ts2-R	ATBCCTTTBTWDGAATGTCCGCC	23	36	76-98		AF233756
Mycosubtilines	Am1-F	CAKCARGTSAAAAATYCGMGG	20	32	1178-1197	<i>B. subtilis</i> ATCC 6633	AAF08795.1
	Tm1-R	CCDASATCAARAADTTATC	20	36	56-76		AAF08796.1 AAF08797.1

Code « IUPAC ADN »: Y= (C or T), M = (A or C), K = (G or T), W= (A or T), D = (G, A or T), S = (G or C), B = (G, T or C), R = (A or G). *La position des amorces est déterminée sur la séquence consensus issues des séquences alignées.

II. Validation des amorces

Les 4 couples d'amorces ainsi définies ont été testés par PCR *in silico* et *in vitro* afin d'évaluer leur spécificité, donc leur capacité à détecter des gènes NRPS.

2.1. Tests PCR *in silico*

Le logiciel AmplifX 1.4.4 a été utilisé pour réaliser des tests PCR *in silico* sur les différents gènes constituant les synthétases de chaque famille de lipopeptides.

Ce test a permis, en outre, de déterminer les différents ORFs amplifiables par chaque couple d'amorces et la capacité de redondance entre elles pour les 4 couples d'amorces.

Le tableau 14 présente les caractéristiques des amorces, la nature et la taille des amplicons obtenus *in silico* sur les différents gènes des synthétases utilisées pour leur définition.

Tableau 14. Tests PCR *in silico* réalisés avec les amorces lipopeptides de *Bacillus* sur les différents ORFs

Amorces	ORFs mycosubtiline synthétase			ORFs surfactine synthétase				ORFs fengycine et plipastatine synthétases				
	<i>mycA</i>	<i>mycB</i>	<i>mycC</i>	<i>srfA-A</i>	<i>srfA-B</i>	<i>srfA-C</i>	<i>srfA-D</i>	<i>fenA/ppsD</i>	<i>fenB/ppsE</i>	<i>fenC/ppsA</i>	<i>fenD/ppsB</i>	<i>fenE/ppsC</i>
Apl-F/Tp1-R	ab	ab	ab	ab	ab	ab	ab	ab /947-959	ab /563-893	ab /929	ab /929	ab /893-1152
Af2-F/Tf1-R	ab	ab	ab	ab	ab	ab	ab	443-452/ ab	455/ ab	ab	ab	ab
As1-F/Ts2-R	ab	ab	ab	419/425/431	419/419/422	ab	ab	ab	ab	ab	ab	ab
Aml-F/Tml-R	ab	419/419/419	416	ab	ab	ab	ab	ab	ab	ab	ab	ab

ab: absence de bande

* bp: base pair

Ces résultats montrent bien la capacité des amorces à hybrider les opérons qui ont servi à leur définition. En général, les amorces détectent des fragments uniquement dans les ORFs des opérons à partir desquels elles ont été définies. Ce qui se traduit par la non-redondance entre ces différents couples.

Les amorces Apl-F/Tp1-R mis au point pour détecter des gènes de la plipastatine synthétase amplifient différents fragments de tailles variables (563 bp, 893 bp, 929 bp et 1152 bp) dans les synthétases, les amplicons des amorces Af2-F/Tf1-R (fengycine), Am1-F/Tm1-R (mycosubtiline et iturine) et As1-F/Ts2-R (surfactine et lichenysine) ont une taille qui oscille entre 416 bp et 455 bp.

Cette variation de taille des amplicons est due essentiellement au positionnement des amorces A et T comme montré dans le tableau 13.

Toutefois, les ORFs *srfA-C*, *srfA-D* et les ORFs *fenD* et *fenE* n'ont pu être détectés respectivement par les amorces As1-F/Ts2-R et les amorces Af2-F/Tf1-R. Ceci est probablement du aux conditions d'utilisation du logiciel utilisé. Il n'apparaît pas exclu que ces amorces puissent reconnaître ces synthétases lors des tests *in vitro*.

2.2. Tests PCR *in vitro*

La température optimale d'hybridation a été déterminée pour chacun des couples d'amorces et les conditions expérimentales pour leur utilisation sont présentées dans le § 2.1.3 du Chapitre Matériels et Méthodes.

Un contrôle de la qualité de la réaction PCR et de l'ADN a été réalisé par l'utilisation des amorces (Prbac1/Prbac2) spécifiques des Eubactéries (Rupf *et al.* 1999). Toutes les amorces ainsi définies ont été testées expérimentalement sur des souches de *Bacillus subtilis*, *B.*

thuringiensis, *B. cereus* et *B. licheniformis*. Les souches de *Lactobacillus plantarum* ATCC 14917 et ATCC 8014 décrites comme ne contenant pas de gènes NRPS par Molenaar *et al.* (2005) ont été utilisées comme témoin négatif. La figure 26 présente l'ensemble des fragments obtenus sur des souches de *Bacillus subtilis* (*B. subtilis* 168, ATCC 21332, ATCC 9943 et ATCC 6633) que nous disposons au sein de notre laboratoire.

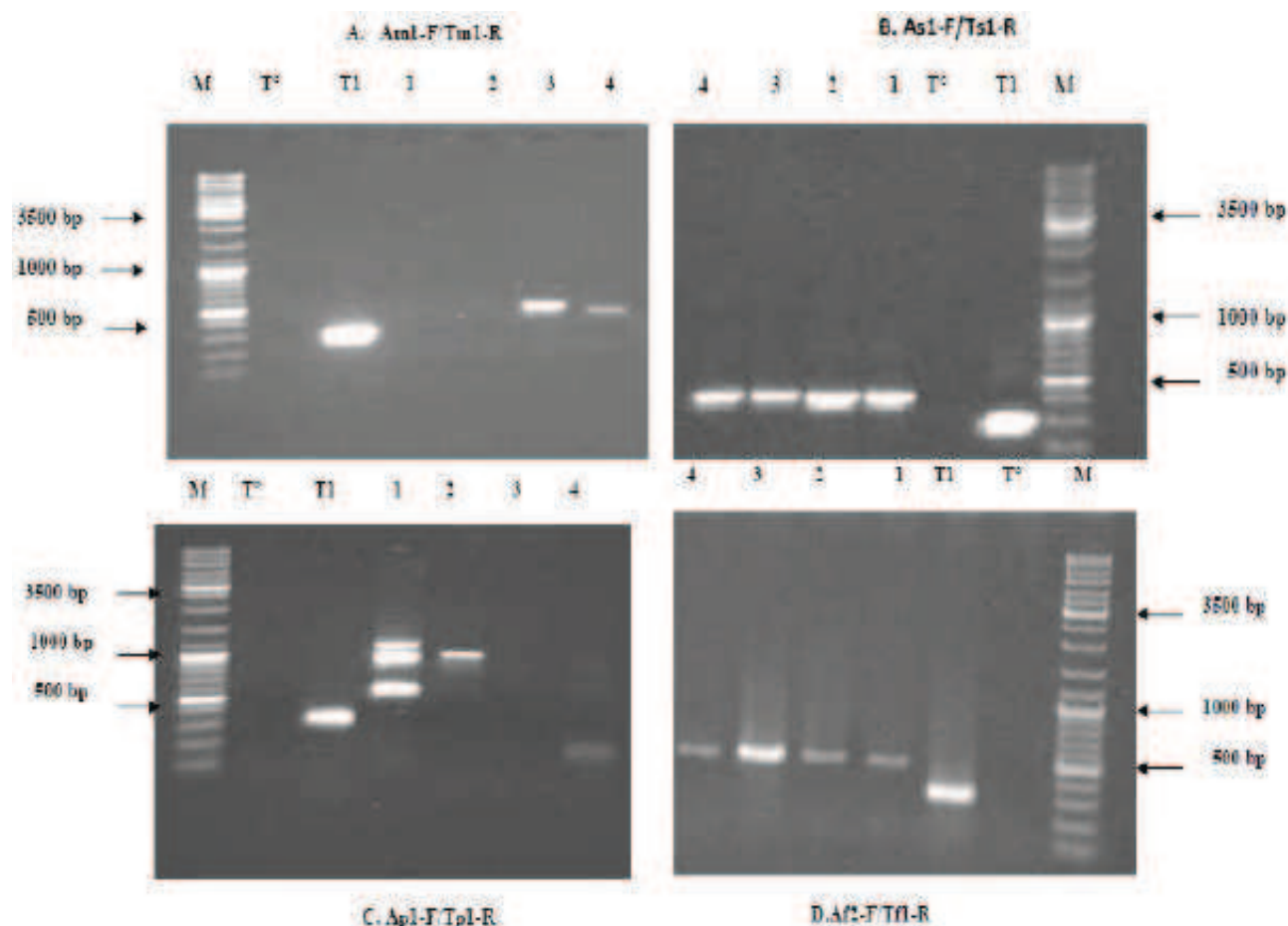


Figure 26. Résultats des réactions PCR *in vitro* sur l'ADN des souches de *Bacillus subtilis* en notre possession (M: O'Gene ruler, T°: échantillon sans ADN, T1: ADN avec amorces PrBac, 1: ADN *B. subtilis* 168, 2: ADN *B. subtilis* ATCC 21332, 3: ADN *B. subtilis* ATCC 6633, 4: ADN *B. subtilis* ATCC 9943)

Les tailles des amplicons obtenus par toutes les amorces correspondent aux tailles obtenues *in silico*. Comme attendu, aucune bande n'a été obtenue avec les ADNs des souches témoin négatif. Les amorces As1-F/Ts2-R et Af2-F/Tf1-R ont amplifié des fragments de tailles attendues respectivement autour de 420 bp et 450 bp (Tableau 14), suggérant la présence de gènes de la surfactine et de la fengycine synthétases dans les souches testées. Les amorces Am1-F/Tm1-R ont détecté des fragments autour de 420 bp chez les souches *B. subtilis* ATCC 9943 et ATCC 6633, signifiant une mise en évidence probable des gènes de synthétase de mycosubtiline ou

d'iturine. De même, les amorces Ap1-F/Tp1-R ont détecté des fragments attendus chez *B. subtilis* 168 et *B. subtilis* ATCC 21332. Comme démontré lors des tests *in silico* (Tableau 14), l'obtention de divers fragments par les amorces Ap1-F/Tp1-R correspond à la différence des tailles des amplicons qu'elles sont susceptibles de détecter.

La très forte intensité observée pour certains amplicons est alors le fait de la superposition de plusieurs fragments de tailles similaires comme dans le cas des amorces As1-F/Ts2-F sur l'ADN de la souche 168. Les fragments ainsi amplifiés ont tous été clonés puis séquencés pour confirmer ou non les résultats de ces tests.

Le tableau 15 présente les résultats des séquences obtenues par Blast N et Blast X sur des tests réalisés sur les ADNs d'autres souches de *Bacillus* (*B. licheniformis* ATCC 14580, *B. thuringiensis* serovar *berliner* 1915 et *B. cereus* LMG 2098) dont le potentiel NRPS est connu ou non.

Tableau 15. Résultats de Blast des séquences de fragments obtenus par PCR avec les amorces dégénérées

Souches	Nom des amorces	Taille amplicons (bp)	N° acc. GenBank des protéines obtenues par BlastX	Match with probable proteins obtained by BlastX	Identité du gène correspondant par BlastN (%)	E. value* de l'identité du gène
<i>B. subtilis</i> ATCC 21332	Af2-F/Tf1-R	449	YP_03393830	Synthétase de la plipastatine de <i>B. subtilis</i> NCIB 3610	93	2e-176
	As1-F/Ts2-R	431	CAB12142	Synthétase de la surfactine A de <i>B. subtilis</i> 168	99	0.00
	Am1-F/Tm1-R	0				
	Ap1-F/Tp1-R	893	CAB13717	Synthétase de la plipastatine de <i>B. subtilis</i> 168	99	0.00
<i>B. subtilis</i> ATCC 9943	Af2-F/Tf1-R	452	YP_001421438	Fengycine C de <i>B. amyloliquefaciens</i> FZB42	97	0.00
	As1-F/Ts2-R	419	YP_001419996	Surfactine B of <i>B. amyloliquefaciens</i> FZB42	97	0.00
	Am1-F/Tm1-R	419	ABY89499	Synthétase IturinA B de <i>B. subtilis</i> MH-25	97	0.00
	Ap1-F/Tp1-R	0				
<i>B. subtilis</i> ATCC 6633	Af2-F/Tf1-R	464	CAB13589	Polyketide synthase type I de <i>B. subtilis</i> 168	88	3e-144
	As1-F/Ts2-R	419	CAA51223	Surfactine A de <i>B. subtilis</i> 168 <i>trpC2</i>	96	0.00
	Am1-F/Tm1-R	419	Q9R9J0	Synthétase mycosubtiline B de <i>B. subtilis</i> 6633	99	0.00
	Ap1-F/Tp1-R	0				
<i>B. subtilis</i> 168	Af2-F/Tf1-R	454	X72672.1	Opéron de la surfactine de <i>B. subtilis</i> 168	99	0.00
	As1-F/Ts2-R	423	CAB12142	Surfactine A de <i>B. subtilis</i> 168	100	0.00
	Am1-F/Tm1-R	0				
	Ap1-F/Tp1-R	959	CAB13717	Plipastatine A de <i>B. subtilis</i> 168	99	0.00
<i>B. licheniformis</i> ATCC 14580	Af2-F/Tf1-R	532	YP_080976	Siderophore 2,3-DHB de <i>B. licheniformis</i> 14580	96	0.00
	As1-F/Ts2-R	428	YP_077642	Lichenysine C de <i>B. licheniformis</i> 14580	99	0.00
	Am1-F/Tm1-R	0				
	Ap1-F/Tp1-R	951	YP_077640.1	Lichenysine A de <i>B. licheniformis</i> 14580	99	0.00
<i>B. thuringiensis</i> <i>berliner</i> 1915	Af2-F/Tf1-R	0				
	As1-F/Ts2-R	416	ACC55454	Protéine NRPS de <i>Acinetobacter baumannii</i>	88	8e-130
	Am1-F/Tm1-R	0				
	Ap1-F/Tp1-R	0				
<i>B. cereus</i> LMG 2098	Af2-F/Tf1-R	0				
	As1-F/Ts2-R	419	CAA51222	Surfactine ORF1 de <i>B. subtilis</i> MH-25	91	8e-155
	Am1-F/Tm1-R	0				
	Ap1-F/Tp1-R	0				

*E. value (Expect Value): une mesure de la valeur représentant la probabilité d'avoir un meilleur résultat que celui affiché.

Toutes les amorces ont détecté des gènes de type NRPS en particulier des lipopeptides, ce qui permet d'affirmer qu'elles ont été sélectives.

Les résultats confinés dans le tableau 15 peuvent être interprétés selon le criblage de trois catégories de souches testées :

- Les souches (*B. subtilis* 168, *B. licheniformis* ATCC 14580 et *B. subtilis* ATCC 6633) dont le génome a été partiellement ou complètement séquencé et dans lequel, la présence de gènes NRPS a déjà été montrée

Comme attendu, les amorces As1-F/Ts2-R ont détecté le gène des synthétases de la surfactine et de la lichenysine respectivement chez *B. subtilis* 168 et *B. licheniformis* ATCC 14580 avec un pourcentage d'identité de 100% et 99%. La détection par le même couple d'amorce de ces deux molécules confirme la capacité des ces amorces à détecter les gènes codant pour des synthétases des deux molécules. Les amorces Ap1-R/Tp1-R ont mis en évidence le gène de la plipastatine de *B. subtilis* 168 avec une similarité de 99% et E. value = 0.00 comme attendu. La mise en évidence de la mycosubtiline de *B. subtilis* ATCC 6633 décrite par Duitman *et al.* (1999) a été détectée par les amorces Am1-F/Tm1-R avec un pourcentage d'identité de 99% et E. value = 0.00.

- Les souches connues pour leur production de molécules NRPS et dont les gènes correspondant sont non-séquencés.

Les amorces As1-F/Ts2-R ont amplifié des fragments de tailles attendues similaire au gène de la synthétase de la surfactine chez *B. subtilis* ATCC 6633 (96% et une E. value = 0.00), *B. subtilis* ATCC 21332 (99%, E. value = 0.00) et chez *B. subtilis* ATCC 9943 (97%, E. value = 0.00), toutes ces souches ayant été décrites comme productrices de la surfactine (Kluge *et al.* 1988; Duitman *et al.* 1999).

La présence de la lichenysine a été aussi détectée chez *B. licheniformis* ATCC 14580 par les amorces Ap1-F/Tp1-R avec un pourcentage d'identité de 99%, E. value = 0.00. Des fragments similaires à 97% (E. value = 0.00) aux gènes de l'iturine et de la fengycine ont été détectés respectivement par les amorces Am1-F/Tm1-R et Af2-F/Tf1-R chez *B. subtilis* ATCC 9943. Ces résultats sont en accord avec l'identification de la souche ATCC 9943 comme co-productrice des trois familles de lipopeptides (Vater *et al.* 2002; Chollet-Imbert *et al.* 2009).

- Les souches non connues pour la production des molécules NRPS détectées

Les amorces Ap1-F/Tp1-R ont permis la détection d'un fragment similaire à 99% avec une E. value = 0.00 chez *B. subtilis* ATCC 21332. Cette souche n'est pourtant pas connue comme productrice de la plipastatine. Ces résultats confirment cependant les travaux de Gancel *et al.* (2009) qui ont démontré récemment la présence d'une molécule de la famille des fengycines chez la souche ATCC 21332.

Les amorces ont été aussi utilisées pour le criblage de deux souches (*B. thuringiensis* serovar *berliner* 1915 et *B. cereus* LMG 2098) inconnues comme arborant un gène NRPS ou produisant une molécule NRPS. Seul le couple d'amorce As1-F/Ts2-R a permis la détection d'un fragment

soit similaire à 88% avec une E. value de $8e-130$ à un gène codant pour une « synthétase de protéine non-ribosomiale » chez *B. thuringiensis berliner* 1915, soit à un gène de la synthétase de surfactine chez *B. cereus* LMG 2098 avec 91% de pourcentage d'identité et une E. value = $8e-155$.

Contrairement aux autres amorces qui ont à chaque fois amplifié des fragments de taille attendue, les amorces Af2-F/Tf1-R ont détecté des fragments non attendus similaires soit à une polykétide synthase de type I chez *B. subtilis* ATCC 6633 et soit à une surfactine synthétase chez *B. subtilis* 168. Ce PKS de type I serait responsable selon Chen *et al.* (2006) de la synthèse de la bacillaene (une molécule hybride NRPS/PKS) chez *B. subtilis* 168. La détection de ce PKS de type I montre que la souche ATCC 6633 est potentiellement productrice de la bacillaene. La détection de ce PKS de type I par les amorces Af2-F/Tf1-R résulte principalement de l'absence du gène de la fengycine synthétase chez *B. subtilis* ATCC 6633. La mise en évidence d'un gène codant pour la surfactine synthétase par ces mêmes amorces chez *B. subtilis* 168 montre aussi que ces amorces ont une affinité plus importante pour les gènes de la surfactine synthétase que pour ceux de la plipastatine de *B. subtilis* 168 (génomé séquencé).

Nous avons alors voulu savoir, pourquoi les amorces Af2-F/Tf1-R censée détecter la présence des gènes de la synthétase de fengycine amplifient *a priori* la surfactine synthétase alors qu'elles auraient pu détecter les gènes de la plipastatine synthétase.

Une étude sur l'hybridabilité des amorces Af2-F/Tf1-R sur les séquences nucléiques de la fengycine, la surfactine et de la plipastatine a été effectuée (Tableau 16). Elle a consisté à déterminer à l'aide du logiciel Vector NTI, la probabilité de l'hybridation de ces amorces sur la zone d'accrochage du fragment de 454 bp amplifié *in vitro* chez *B. subtilis* 168 (similaire à 99% à la surfactine) et les gènes de fengycine, surfactine et de la plipastatine synthétases.

Tableau 16. Détermination des pourcentages d'accrochage des amorces Af2-F/Tf1-R sur les gènes de la fengycine, la surfactine et de la plipastatine

	Gènes des synthétases	Zone accrochage amorces (%)	
		Af2-F	Tf1-R
Séquence obtenue <i>in vitro</i> par les amorces Af2-F/Tf1-R sur <i>B. subtilis</i> 168	<i>fen A</i>	69.0	85.0
	<i>fen B</i>	80.0	76.2
	<i>fen C</i>	78.3	83.3
	<i>fen D</i>	90.0	81.0
	<i>fen E</i>	78.3	85.7
	<i>srfA-A</i>	69.6	81.0
	<i>srfA-B</i>	73.9	85.0
	<i>srfA-C</i>	73.9	69.6
	<i>ppsA</i>	54.5	58.3
	<i>ppsB</i>	70.8	57.7
	<i>ppsC</i>	40.0	60.0
	<i>ppsD</i>	57.1	52.6
	<i>ppsE</i>	52.2	65.0

L'analyse du tableau comparatif montre qu'en dehors de la fengycine, les amorces Af2-F et Tf1-R hybrident respectivement avec une moyenne de 72,46 et 78,53 le gène de la surfactine synthétase, contre en moyenne 54,92 et 58,72 le gène de la plipastatine synthétase. Cela explique le fait que ces amorces ont détecté plutôt un fragment similaire à la surfactine synthétase qu'un fragment similaire à la plipastatine synthétase.

Tout se passe comme si la fengycine de *B. subtilis* F29-3 est plus proche de la surfactine de *B. subtilis* 168 que de sa plipastatine. Ceci confirme bien la différence de 25% entre les opérons de fengycine et de plipastatine (Tsuey-pin *et al.* 1999), et le fait que nous avons choisi de définir deux couples d'amorces différentes pour hybrider ces deux molécules appartenant à la même famille protéique.

Conclusions et Perspectives

La technique consistant à définir des amorces à partir des domaines d'adénylation et de thiolation par l'alignement de leur séquence nucléique en tenant compte des différences au niveau des opérons est une approche originale permettant de mettre en évidence la possible expression de ces molécules.

L'application de cette méthode aux lipopeptides synthétases a permis la mise en évidence de gènes correspondants chez des souches de *Bacillus*.

Ainsi pour pouvoir détecter d'autres gènes NRPS sur d'autres genres, il faudrait tenir compte de l'organisation des synthétases à l'intérieur de chaque genre en appliquant la technique comme dans le cas des *Bacillus*.

L'un des inconvénients de l'utilisation des amorces définies à partir des opérons lipopeptides est qu'elles peuvent ne pas mettre en évidence la présence d'autres gènes NRPS lorsque le génome bactérien contient un gène impliqué dans la synthèse d'un lipopeptide. Cette défaillance pourrait être palliée lors des tests de criblage avec plusieurs couples d'amorces ou en multiplex PCR.

Valorisation

- ✓ **A. Tapi, M. Chollet-Imbert, B. Scherens, Ph. Jacques**, 5th International conference on Gram-Positive Microorganisms/15th International Conference on Bacilli, San-Diego, California. June 14-18, 2009.

- ✓ **A. Tapi, M. Chollet-Imbert, B. Scherens, Ph. Jacques** (2009) New approach for the detection of non ribosomal peptide synthetase genes in *Bacillus* strains by Polymerase Chain Reaction. App Microbiol Biotechnol. DOI 10.1007/s00253-009-2176-4 (annexes).

- ✓ **Jacques P, Froidevaux R, Chollet-Imbert M, Tapi A, Caboche S, Pupin M, Kuchеров G, Dhulster P, Le Flem G, Vercaigne-Marko D, Leclère V**. Novel approaches to isolate antimicrobial peptides using bioinformatics and molecular tools. Second international symposium on antimicrobial peptides-food, veterinary, medical and novel applications. St Malo (France), communications, 17-19 June 2009.

Chapitre 7 Identification de souches du genre *Bacillus* productrices de lipopeptides d'origine non ribosomiale isolées de deux produits fermentés africains : le Soumbala et le Bikalga

Le Soumbala et le Bikalga sont des condiments fermentés fabriqués respectivement à partir de *Parkia biglobosa* communément appelé néré et d'*Hibiscus sabdariffa* (Oseille de Guinée ou roselle). Ces aliments sont décrits dans le § 2.3.4 du Chapitre Matériels et Méthodes.

Toutes les souches ont été isolées de ces aliments fermentés par le Dr. Aly Savadogo de l'Université de Ouagadougou (Burkina-Faso). En collaboration avec celui-ci, nous nous sommes intéressés exclusivement à celles appartenant au genre *Bacillus*. Le but de cette étude est de sélectionner les souches capables de synthétiser des lipopeptides non ribosomiaux.

Dans ce chapitre, nous présentons la combinaison des techniques classiques d'identification biochimiques et moléculaires pour, dans un premier temps caractériser les souches de *Bacillus* puis dans un deuxième temps, mettre en application les techniques de détection de gènes NRPS avec les amorces développées dans les chapitres précédents associées aux techniques de Maldi-TOF pour déterminer les souches possédant la capacité de synthétiser des lipopeptides.

I. Identification des souches issues des produits fermentés

Toutes les souches ont été isolées sur des milieux spécifiques aux bactéries lactiques et non lactiques en fonction de leur température optimale de croissance, puis elles ont été caractérisées par des tests d'identification biochimique et moléculaire.

1.1. Isolement des souches

Une centaine de souches de bactéries lactiques et non lactiques ont été isolées des aliments fermentés (Soumbala et Bikalga).

Des codes spécifiques ont été attribués et des échantillonnages ont été effectués comme décrits dans le § 2.3.4 du Chapitre Matériels et Méthodes.

Les souches isolées des différents échantillons sont présentées dans les tableaux 17, 18 et 19 en fonction du milieu de culture, de la température optimale de croissance et du condiment de provenance de la souche.

Le tableau 17 présente les souches isolées à partir de 6 échantillons de Soumbala (S), sur Bouillon Nutrive (BN) à 30°C et 37°C.

Tableau 17. Souches isolées sur BN (à 30°C et 37°C) à partir du Soumbala

Echantillons	Souches à 30°C	Souches à 37°C
S1	0S1B012, 0S1B023, 0S1B022, 0S1B021	7S1B012, 7S1B011, 7S1B032, 7S1B031
S2	0S2B051, 0S2B021, 0S2B052 0S2B022, 0S2B031	7S2B041, 7S2B042
S3	0S3B011, 0S3B021, 0S3B042 0S3B061, 0S3B032, 0S3B041	7S3B012, 7S3B011, 7S3B041
S4	0S4B031, 0S4B022, 0S4B032, 0S4B021	7S4B012, 7S4B011, 7S4B021, 7S4B022, 7S4B041
S5	0S5B041, 0S5B012, 0S5B011, 0S5B042	7S5B041, 7S5B031
S6	*0S6B021	7S6B021, 7S6B011

*Photographie figure 27

42 souches différentes ont été isolées des différents échantillons selon leur température de croissance. La figure 27 montre une photo de la souche 0S6B021 isolée du Soumbala.



Figure 27. Photographie de la souche **0S6B021**

Le tableau 18 présente les souches qui ont été isolées sur milieu BN à 30°C et 37°C des échantillons issus de Bikalga (B).

Tableau 18. Souches isolées sur BN (à 30°C et 37°C) à partir du Bikalga (B)

Echantillons	Souches à 30°C	Souches à 37°C
B7	0B7B011, 0B7B032, 0B7B031, 0B7B041, 0B7B042	7B7B031, 7B7B032, 7B7B021, 7B7B011, 7B7B041
B8	0B8B011, 0B8B012, 0B8B021	7B8B011
B9	0B9B011	7B9B021
B10	0B10B011, 0B10B022, 0B10B031, 0B10B021	7B10B011, 7B10B021, 7B10B031

23 souches différentes ont été isolées des échantillons de Bikalga selon leur température de croissance.

La recherche de souches dans ces aliments a été aussi réalisée à partir du milieu MRS à 30°C et 37°C (Tableau 19).

Tableau 19. Souches isolées du Bikalga (B) et du Soumbala (S) sur le milieu MRS (30°C et 37°C)

Echantillons	Souches à 30°C	Souches à 37°C
S1	abs	abs
S2	abs	7S2M021, 7S2M022
S3	abs	abs
S4	0S4M021	7S4M021, 7S4M022
S5	abs	abs
S6	abs	7S6M041
B7	abs	abs
B8	0B8M011, 0B8M012	abs
B9	0B9M012, 0B9M011	abs
B10	0B10M011	7B10M021, 7B10M022

abs : absence de souches

De nombreuses souches ont été isolées sur les deux types de milieu d'incubation (MRS et BN) dans les deux types d'aliments. Bien que des redondances à partir des 2 milieux et des 2 températures d'incubation (malgré la sélection basée sur des critères morphologiques et culturels) ne puissent être écartées, ces résultats montrent la diversité de la population bactérienne dans ces aliments fermentés, ce qui souligne l'intérêt de les étudier.

La répartition des souches en fonction du milieu de culture, de la température de croissance et des aliments est présentée dans le tableau 20.

Tableau 20. Distribution du nombre de souches isolées en fonction de la température et des aliments

	BN		MRS	
	30°C	37°C	30°C	37°C
Soumbala	24	18	1	5
Bikalga	13	10	5	7

Au total 83 souches ont été isolées des ces échantillons selon le milieu d'incubation et la température de croissance, puis de l'aspect visuel des colonies. On remarque qu'il y a plus de colonies isolées sur bouillon nutritif que sur milieu MRS ce qui suggère que la flore lactique est peu présente dans ces condiments.

1.2. Identification des souches isolées

Pour affiner l'identification des souches et sélectionner celles ayant un profil *Bacillus*, des tests de biochimie classique et moléculaire ont été effectués sur celles ayant exprimées une ou plusieurs activités biologiques. Ces dernières sont en effet susceptibles de présenter un intérêt lié à leur potentiel NRPS.

1.2.1. Tests biologiques et physico-chimiques

Le potentiel biologique des souches isolées a été analysé à travers l'activité antibactérienne, antifongique, antilevure et hémolytique selon les techniques décrites dans le § 2.3.3 du Chapitre Matériels et Méthodes.

- Activités antibactériennes

La recherche d'activités antibactériennes a été effectuée à l'aide de la bactérie cible du genre *Micrococcus* par la technique des puits. La présence de certaines *Micrococcus* dans le sang provoque une bactériémie et une infection grave. On la retrouve parfois dans de l'eau et dans certains aliments comme le fromage. Les résultats obtenus sont présentés dans le tableau 21 suivant.

- Activités antifongiques

Le potentiel antifongique des souches a été recherché par des tests de rapprochement cellules-champignons ou dits de compétition cellulaire. Quatre souches de moisissures (*Aspergillus niger*, *Penicillium roquefortii*, *Mucor rouxii* et *Rhizopus nigricans*) et une souche de levure (*Saccharomyces cerevisiae*) décrites dans le § 1.2.3 du Chapitre Matériels et Méthodes, ont été utilisées pour ces tests. Ces moisissures ont été choisies d'une part, parce qu'elles sont connues pour leur innocuité vis-à-vis du genre humain (germe de classe 1) mais aussi parce qu'elles sont capables d'infecter des plantes ou que certaines espèces de même genre sont proches de germes pathogènes comme *Aspergillus fumigatus*, responsable d'aspergillose chez les humains (Kac *et al.* 1995).

Les résultats sont confinés dans le tableau 21, présentant pour chacun des germes cibles, les souches ayant réagi positivement en fonction de leurs milieux d'isolement.

- **Activités hémolytiques**

Pour vérifier l'activité hémolytique éventuelle de ces souches, des tests sur géloses au sang ont également été réalisés (Tableau 21). En effet, certaines souches du genre *Bacillus* sont décrites comme synthétisant des lipopeptides (comme la lichenysine et la surfactine) qui possèdent un potentiel hémolytique (Deleu *et al.* 2003).

Tableau 21. Activités antifongiques et antimicrobiennes des souches isolées de condiments africains

Tests antimicrobiens	Tests antifongiques			Tests hémolytiques	Tests antilevures
<i>Micrococcus luteus</i>	<i>Penicillium roquefortii</i>	<i>Mucor rouxii</i>	<i>Rhizopus nigricans</i>	Gélose au sang	<i>Saccharomyces cerevisiae</i>
Code (*NA)	Code (NA)	Code (NA)	Code (NA)	Codes (NA)	Code (NA)
0S5B011 0S1B012 (2) 7S6B021(4) 7S6M041 0S2B052 (6) 0B7B031 (7) 0B7B032 (8) 0B8B021 (9) 0B9B011 (10) 0B10B021(11) 0S6B021 (12) 0B10B022 (13) 7B8B011 7S4B011 0B10B031 0B10M011 7S1B031(15)	0B10B021 (21) 0B9B011 (10) 0B10B022 (13)	0B10B021 (21) 0B10B022 (13)	7B7B041 7S6B021 (4) 0S1B023 7S6B011 (16) 0B10B021 (21) 0B9B011 (10) 0S4B021 (5) 0B7B032 (8) 0B10M011 (14) 0S6B021 (12) 0S3B032 0S5B011 7B10B021 (19) 0B10B022 (13) 0S1B012 0B8M011 (18) 7B7B021 0B9M011	7S2B041 0S1B022 7S1B032 (21) 0B7B042 0B9B021 0B7B041 7S2B042 0S1B021 0S2B052 (6) 7B7B021 (19) 7S1B031 (15) 0S1B021 0B8M012 7B7B021 7S1B031 (20) 0B8M012 (22) 0B8M011 (18)	0S4M021 (1) 7S5B041 7S3B041 0S4B021 (5)

*NA: numéro affecté

Parmi toutes les souches isolées des échantillons de Soumbala et Bikalga, quelque soit le milieu d'incubation, nous en avons sélectionné 22 qui ont exprimé au moins une activité biologique pouvant être un indicateur de la présence de lipopeptides NRPS. Ces souches ont alors été numérotées de 1 à 22 (Tableau 21).

17 souches ont exprimé une activité contre *Micrococcus luteus*. Aucune des souches testées n'a été capable d'inhiber la croissance d'*Aspergillus niger*. Certaines d'entre elles comme 0B10B022 (souche 13) ont exprimé plusieurs activités antifongiques (contre *P. roquefortii*, *M. rouxii* et *R. nigricans*). Ces tests ne permettent pas de savoir si ces activités antifongiques sont la conséquence d'une ou de plusieurs molécules synthétisées. D'autres souches ont exprimé une activité hémolytique et/ou anti-levure.

Les souches numérotées de 1 à 22 ont fait l'objet d'une caractérisation approfondie.

1.2.2. Caractéristiques morphologiques et biochimiques des souches

- Caractéristiques morphologiques

Les caractéristiques morphologiques (diamètre des colonies, couleur et aspect) des 22 souches ont été répertoriées. Le Tableau 22 présente les résultats obtenus.

Tableau 22. Caractéristiques morphologiques des 22 colonies de souches sélectionnées

Souches	Numéro	Couleur	Taille (mm)	Forme
0S4M021	1	blanche	1	ronde
0S1B012	2	blanche	0,5	ronde
7S5B041	3	blanche	2,5	ronde
7S6B021	4	blanche	2	ronde
0S4B021	5	blanche	2,2	dentelée
0S2B052	6	blanche	0,6	ronde
0B7B031	7	blanche	3	ronde
0B7B032	8	blanche	1	dentelée
0B8B021	9	blanche	2	dentelée
0B9B011	10	blanche	3	dentelée
0B10B021	11	blanche	3	ronde
0S6B021	12	blanche	2	dentelée
0B10B022	13	blanche	1	dentelée
0B10M011	14	blanche	1,5	ronde
7S1B031	15	blanche	2	ronde
7S6B011	16	blanche	1	ronde
7B10B021	17	blanche	0,2	ronde
0B8M011	18	blanche	1,5	ronde
7B7B021	19	blanche	2	dentelée
7B10B031	20	blanche	2	dentelée
7S1B032	21	blanche	1	ronde
0B8M012	22	blanche	3	ronde

Les colonies volumineuses, plates et opaques, à surface chagrinée et à bord légèrement dentelées montrent des caractéristiques morphologiques semblables à celles appartenant au genre *Bacillus*. Une coloration de Gram et un test de la production de catalase ont été réalisés sur les 22 souches. Le tableau 23 récapitule les résultats de tous les tests réalisés. Ces souches sont majoritairement des bacilles catalase +, suggérant une appartenance au genre *Bacillus*. Ceci semble confirmer les travaux de Parkouda *et al.* (2008) qui ont montré que la majeure partie des souches retrouvées dans le Soumbala et le Bikalga sont de type *Bacillus*.

Tableau 23. Divers caractéristiques des souches d'intérêt

Souches	N°	Gram	Catalase	Morph	Antibact	Antifongique	Hémolytique
0S4M021	1	+	+	bacille	-	+/-	-
0S1B012	2	+	+	coque	+	+	+
7S5B041	3	-	+	coque	-	+	-
7S6B021	4	+	+	bacille	+	+	+
0S4B021	5	-	+	bacille	-	+	-
0S2B052	6	+	+	bacille	+	+/-	+
0B7B031	7	-	+	bacille	+	-	+
0B7B032	8	+	+	bacille	+	+	-
0B8B021	9	+	+	bacille	+	-	+
0B9B011	10	-	+	bacille	+	+	-
0B10B021	11	+/-	+	bacille	+	+	-
0S6B021	12	+	+	bacille	+	+	+
0B10B022	13	+	+	bacille	+	+	-
0B10M011	14	+	+	bacille	+	+	-
7S1B031	15	-	+	bacille	+	+/-	-
7S6B011	16	+	+	bacille	-	+	+
7B10B021	17	+	+	bacille	-	+	+/-
0B8M011	18	-	+	bacille	-	+	+
7B7B021	19	-	+	coque	-	+	+
7B10B031	20	+	+	bacille	-	+	+
7S1B032	21	+/-	+	bacille	-	+	+
0B8M012	22	-	+	bacille	-	+/-	+

+ : positif, - : négatif

1.2.3. Tests moléculaires d'identification

Des analyses moléculaires à l'aide d'amorces spécifiques du genre *Bacillus*, du groupe *Bacillus cereus* et du groupe *Bacillus subtilis* ont permis de compléter l'identification des 22 souches.

Les résultats sont dits positifs lorsque les fragments obtenus par PCR sont similaires (taille et intensité des bandes) à ceux attendus comme présenté dans le tableau 6.

Les amorces B-K1-F/ B-K1-R1 (identification du gène ARN 16s de *Bacillus*, *Paenibacillus*, *Brevibacillus*) utilisées dans les conditions prédéfinies dans le § 2.3.1 du Chapitre Matériels et Méthodes ont permis de vérifier si les souches appartenaient au genre *Bacillus*.

La figure 28 montre un exemple de résultats obtenus par PCR.

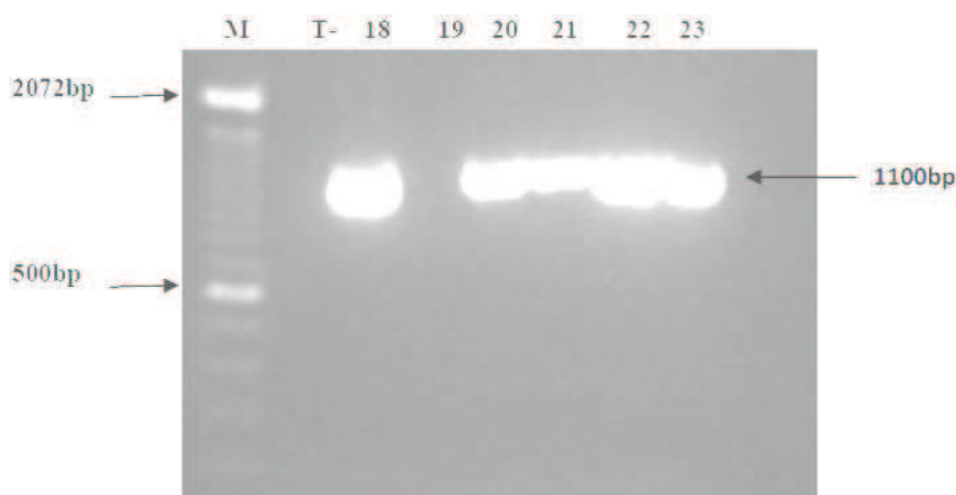


Figure 28. Résultats PCR sur gel d'agarose de l'ADN de souches testées avec les amorces B-K1-F/ B-K1-R1 (M = Trak-It 100 DNA Ladder, T- : témoin négatif avec DNA sans amorces, les puits 18, 19, 20, 21 et 22 sont des souches d'intérêt testées), le puits 23 une souche témoin de *Bacillus subtilis*.

Ce couple d'amorces a conduit à l'amplification de fragments de tailles attendues (environ 1100 bp) chez 15 des 22 souches testées. Il s'agit des souches 1, 3, 5, 6, 9, 10, 11, 13, 14, 15, 16, 18, 20, 21, 22. Le couple d'amorces Bsub5F/Bsub3R (spécifique du groupe *Bacillus subtilis*) a généré des fragments de tailles attendues (595 bp) avec 8 souches sur les 22 souches d'intérêt. Il s'agit des souches 2, 3, 6, 9, 18, 20, 21, 22. Seule la souche 1 a répondu positivement au couple d'amorce ba1F/ba1R (pour le groupe des *Bacillus cereus*) sur les 22 souches testées. Cependant aucun profil génétique n'a pu être établi pour les souches 4 et 17.

Le tableau 24 résume les résultats des tests PCR obtenus sur les souches testées.

Tableau 24. Profil génotypique pour l'identification moléculaire des souches criblées

Amorces	Souches positives	Genre ou espèce
B-K1/F	1, 3, 5, 6, 9, 10, 11, 13, 14	<i>Bacillus</i> (<i>subtilis</i> , <i>cereus</i> , <i>thuringiensis</i> , <i>licheniformis</i>)
B-K1/R1	15, 16, 18, 20, 21, 22	
Bsub5F	2, 3, 6, 9, 18, 20, 21, 22	Groupe <i>subtilis</i> group (<i>subtilis</i> , <i>licheniformis</i> , <i>amyloliquefaciens</i> , <i>pumilus</i>)
Bsub3R		
Ba1F	1	Groupe <i>Bacillus cereus</i> group (<i>thuringiensis</i> , <i>anthracis</i> et <i>cereus</i>)
Ba1R		

En couleur les fragments qui ont été séquencés

Nous avons séquencé les fragments obtenus pour les souches 6, 11, 13, 16 et 21 avec les amorces spécifiques BK1-F/BK1-R1 pour vérifier s'ils correspondaient bien à des fragments similaires à ceux appartenant au genre *Bacillus*.

Les figures 29 et 30 présentent les séquences respectives des fragments d'environ 1100 bp amplifiés avec les amorces BK1-F/BK1-R1 chez les souches 13 (S13) et 21 (S 21).

BK1-F

TCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACG
 GCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGC
 AACGCCGCGTGAGTGATGAAGGTTTTTCGGATCGTAAACCTCTGTTGTTAGGGAAGAACAAGTAC
 CGTTCGAATAGGGCGGTACCTTGACGGTACCTAACAGAAAGCCACGGCTAACTACGTGCCAGC
 AGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCAGGC
 GGTTTCTTAAGTCTGATGTGAAAGCCCCGGCTCAACCGGGGAGGGTCATTGGAAACTGGGGAA
 CTTGAGTGCAGAAGAGGAGAGTGGAATTCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGG
 AACACCAAGTGGCGAAGGCGACTCTCTGGTCTGTAACCTGACGCTGAGGCGCGAAAGCGTGGGGAG
 CGAACAGGATTAGATACCTTGGTAGTCCACGCGTAAACGATGAGTGCTAAGTGTTAGAGGGTT
 TCCGCCCTTTAGTGCTGCAGCAAACGCATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACT
 GAAACTCAAAGGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTGCAAGCAA
 CGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAACCTAGAGATAGGGCTTCCCTTCGG
 GGGCAGAGTGACAGGTGGTGCATGGTTGTCTGCTAGCTCGTGTCTGTGAGATGTTTGGGTAAAGTC
 CCGCAACGAGCGCAACCTTGATCTTAGTTGCCAGCATTAGTTGGGCACTTAAGGTGACTGC
 CGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCCTTATGACCTGGGCTA
 CACACGTGCTACAATGGGCAGAACAAAGGGCAGCGAAACCGCGAGGCTAAGCCAATCCCACAAA
 TCTGTTCTCAGTTCGGATCGCAGTCTGCAACTCGACTGCGGAAGCTGGAATCGCTAGTAATCGC
 GGATCAGCATGCCGCGGTGAATAG

BK1-R1

Figure 29. Séquence du fragment de 1100 bp amplifié par les amorces BK1-F/BK1-R1 de la souche S 13

BK1 - F

TCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACG
 GCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGC
 AACGCCGCGTGAGTGATGAAGGTTTTTCGGATCGTAAAGCTCTGTTGTTAGGGAAGAACAAGTAC
 CGTTCGAATAGGGCGGTACCTTGACGGTACCTAACAGAAAGCCACGGCTAACTACGTGCCAGC
 AGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGCAGGC
 GGTTTCTTAAGTCTGATGTGAAAGCCCCGGCTCAACCGGGGAGGGTCATTGGAAACTGGGGAA
 CTTGAGTGCAGAAGAGGAGAGTGGAATTCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGG
 AACACCAAGTGGCGAAGGCGACTCTCTGGTCTGTAACCTGACGCTGAGGAGCGAAAGCGTGGGGAG
 CGAACAGGATTAGATACCTTGGTAGTCCACGCGCTAAACGATGAGTGCTAAGTGTTAGGGGGTT
 TCCGCCCTTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACT
 GAAACTCAAAGGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTGCAAGCAA
 CGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAATCCTAGAGATAGGACGTCCCTTCGG
 GGGCAGAGTGACAGGTGGTGCATGGTTGTCTGCTAGCTCGTGTCTGTGAGATGTTGGGTTAAGTCC
 CGCAACGAGCGCAACCTTGATCTTAGTTGCCAGCATTAGTTGGGCACTTAAGGTGACTGCC
 GGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCCTTATGACCTGGGCTAC
 ACACGTGCTACAATGGACAGAACAAAGGGCAGCGAAACCGCGAGGTTAAGCCAATCCCACAAAT
 CTGTTCTCAGTTCGGATCGCAGTCTGCAACTCGACTGCGTGAAGCTGGAATCGCTAGTAATCGC
 GGATCAGCATGCCGCGGTGAATAG

BK1-R1

Figure 30. Séquence du fragment de 1100 bp amplifié par les amorces BK1-F/BK1-R1 de la souche S 21

Le fragment amplifié chez la souche 13 (figure 29) mesure en réalité 1114 bp (similaire à la taille du fragment attendu pour les amorces BK1-F/BK1-R1). L'analyse de ce fragment par Blast N montre qu'il est similaire (99% avec une E. value = 0.00) à un fragment de *Bacillus licheniformis*.

Celui détecté par les mêmes amorces chez la souche S 21 (figure 30) mesure 1113 bp, est similaire à la taille du fragment attendu et identique à 100% (E. value = 0.00) à un fragment de *Bacillus subtilis*.

Les autres souches (6, 11 et 16) ont été identifiées après le séquençage comme appartenant respectivement à des souches de *B. subtilis* pour la S 6 et S 16 alors que la S 11 correspondrait à une *B. licheniformis*.

L'identification des fragments détectés par PCR après le séquençage confirme donc que toutes les souches testées correspondent bien au genre *Bacillus* excepté les souches 4 et 17.

Conclusions sur les tests d'identification

Le croisement des deux types d'identification (biochimique et moléculaire) des 22 souches a permis de sélectionner certaines souches (5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 20 et 21) selon des critères spécifiques aux genres *Bacillus* (bacilles Gram+, catalase +, culture en aérobiose sur milieu ordinaire profil Api 50 de *Bacillus* (Dr Savadogo). Celles-ci ont été retenues pour la suite de l'étude.

Ces souches appartiennent à plusieurs espèces telles que *Bacillus subtilis*, *B. cereus* et *B. licheniformis*. Ces espèces sont bien connues pour leur production de molécules non ribosomiales en particulier les lipopeptides. Notre attention s'est donc portée sur ces 12 souches pour rechercher celles susceptibles de produire des lipopeptides non ribosomiaux.

II. Recherche de lipopeptides non ribosomiaux

Trois techniques ont été mises en œuvre pour rechercher les lipopeptides: la mesure de la tension de surface des surnageants de culture, une analyse MALDI-TOF sur les surnageants de culture et sur colonies enfin des analyses par PCR pour détecter des gènes NRPS. Les analyses biochimiques sont appliquées sur des surnageants de culture et sur des colonies obtenues sur milieu solide, les analyses par PCR ont été effectuées à l'aide des amorces dégénérées spécifiques NRPS mis au point dans les chapitres précédents.

2.1. Analyses biochimiques

2.1.1. Tension de surface des surnageants de culture

La tension de surface peut se définir comme étant la capacité d'adhérence à l'interface entre des liquides non miscibles. Ces liquides vont subir deux forces: une attraction vers l'intérieur "eux-mêmes" et une répulsion par rapport aux molécules. On parle de forces d'attraction intra-liquide et de forces de répulsion inter-liquide. La tension de surface est aussi influencée par la pesanteur.

La production de lipopeptides par des bactéries en milieu liquide va provoquer une baisse de la tension de surface de l'eau.

Nous avons déterminé dans les surnageants de culture des souches isolées, la tension de surface par la méthode de Du Nouy décrite dans le § 2.4.1 du Chapitre Matériels et Méthodes par rapport au milieu d'incubation.

La figure 31 présente un graphique de la mesure des tensions de surface obtenues pour les souches testées.

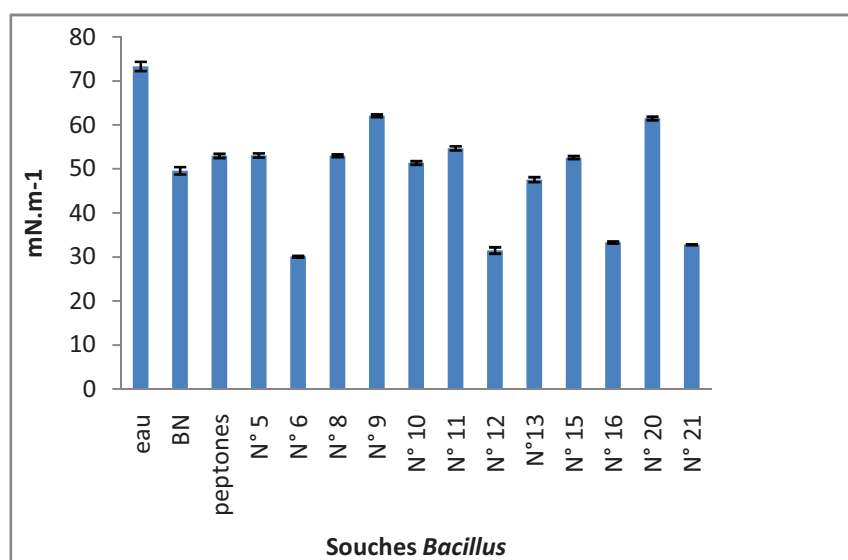


Figure 31. Détermination de la tension de surface des souches issues des condiments fermentés

Les souches 6, 12, 16 et 21 ont une tension de surface faible ($< 33 \text{ mN.m}^{-1}$) par rapport à celle du milieu BN ($49,5 \text{ mN.m}^{-1}$), ce qui traduit qu'elles produisent des métabolites ayant une activité biosurfactante élevée contrairement aux autres souches testées qui ont exprimé une faible activité. Les souches 6, 16 et 21 ont été identifiées comme étant du genre *B. subtilis* et la souche 12 comme de l'espèce *licheniformis*. A ces dernières, pour la suite de l'étude, ont été associées les souches 11 et 13 identifiées comme *Bacillus licheniformis* et ayant une tension de surface proche de celle du milieu BN.

2.1.2. Recherche de lipopeptides par Maldi-TOF

Des analyses de détection de lipopeptides par Maldi-TOF ont été effectuées sur des colonies selon la technique décrite dans le § 2.4.3 du Chapitre Matériels et Méthodes. Les surnageants ont été obtenus à partir de filtrats de lipopeptides à l'aide de colonnes Altech C18 comme décrit dans le § 2.4.2 du Chapitre Matériels et Méthodes.

Les figures 32A et 32B présente les spectres des masses obtenus avec la souche 6 respectivement à partir des milieux solide et liquide.

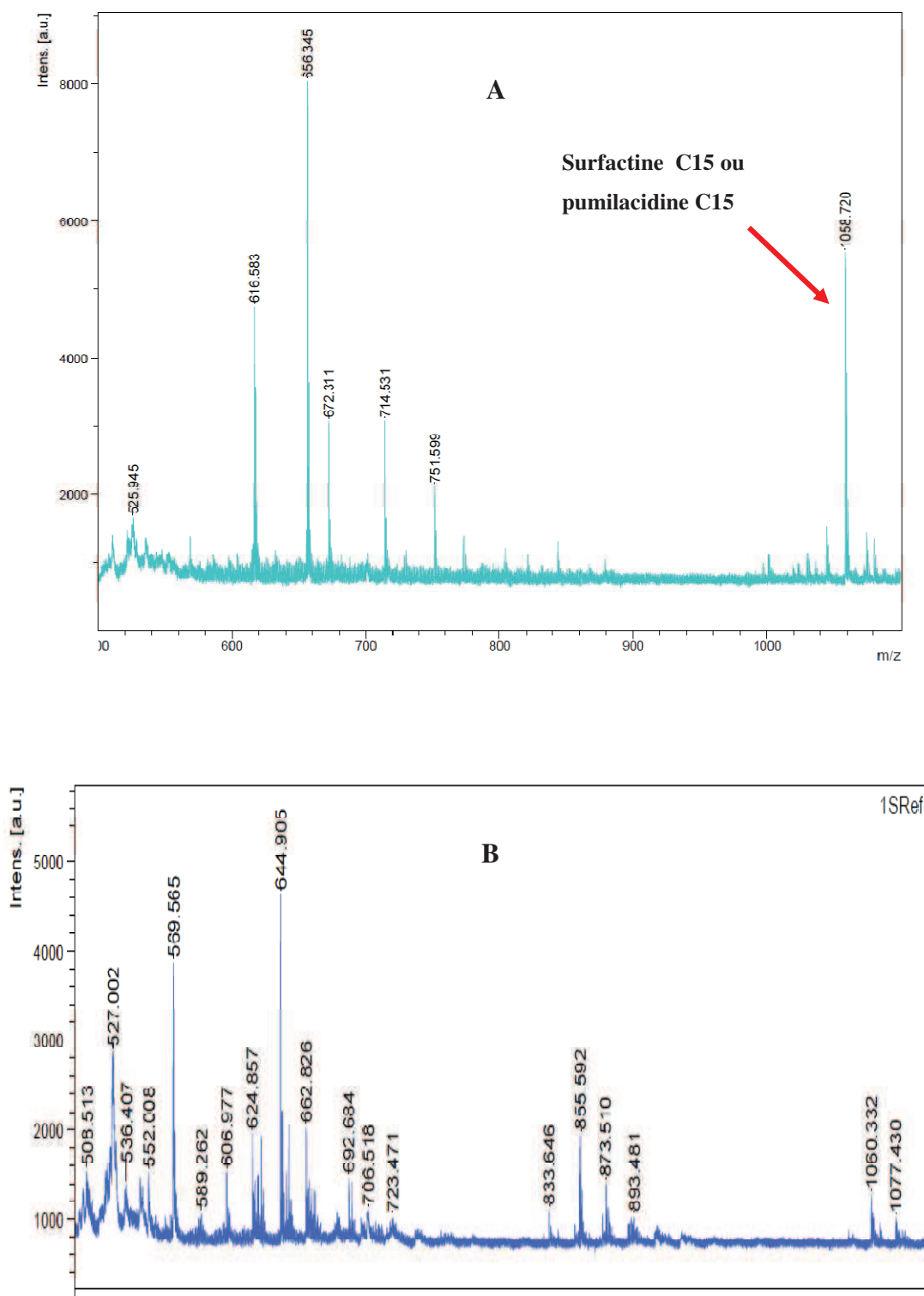


Figure 32. Spectres obtenus à partir de la souche S 6 (**A.** colonie sur milieu solide GN, **B.** surnagent sur BN)

Les spectres des 6 souches ainsi obtenus ont été analysés à la fois pour les surnageants de culture et pour les colonies par rapport aux masses connues de lipopeptides de *Bacillus*. En effet, les masses de lipopeptides de *Bacillus* sont rangées en trois classes distinctes: $m/z = 850-950$ correspondant aux kurstakines, $m/z = 1000-1100$ correspondant aux surfactines et iturines, $m/z = 1450-1550$ correspondant aux fengycines (Price *et al.* 2007).

En général, les pics détectés en milieu solide sont différents de ceux du milieu liquide pour la même souche. On note la présence de la surfactine C15 ou de pumilacidine C15 ($m/z = 1058$) chez les souches 6 et 12. Cette molécule a été détectée uniquement sur milieu solide ce qui est probablement en relation avec son adhérence à la paroi cellulaire. D'autres spectres de masses ne correspondant pas à des lipopeptides connus ont été aussi mis en évidence.

On remarque des spectres de masses proches de celles des kurstakines (877, 892 et 906) dans les souches 6, 12, 13 et 16 (figure 32 B).

Conclusion sur les techniques d'analyses biochimiques

La combinaison des tests de mesure de la tension de surface et de l'analyse Maldi-TOF confirme la présence de la surfactine dans les souches 6 et 12. Toutefois, les analyses Maldi-TOF n'ont pu montrer la présence de lipopeptides dans les souches 11, 13, 16 et 21 contrairement à la faible tension de surface obtenue dans ces dernières.

Des analyses au niveau moléculaire ont donc été effectuées pour confirmer la présence de gènes connus ou séquencés pour ces souches.

2.2. Recherche de gènes de lipopeptides synthétases

Des tests pour l'identification de peptides bioactifs produits par les 6 souches d'intérêt ont été réalisés à l'aide des amorces dégénérées spécifiques du genre *Bacillus* (Tapi *et al.* 2009). Trois couples d'amorces (As1-F/Ts2-R, Am1-F/Tm1-F et Af2-F/Tf1-R) ont été utilisés.

Les résultats sont dits positifs lorsque la taille et l'intensité des bandes sont similaires à celles décrites pour chacun des couples d'amorces.

Le résumé des tailles de bandes obtenues pour chaque souche pour les 3 couples d'amorces est représenté dans le tableau 25.

Tableau 25. Taille des bandes amplifiées par couple d'amorce

Souches	As1-F/Ts2-R (bp)	Am1-F/Tm1-F(bp)	Af2-F/Tf1-(bp)
S6	420	450	450
S11	420	pb	450
S12	420	400	pb
S13	420	*pb	450
S16	420	400	pb
S21	420	450	pb

*pb : pas de bandes

Uniquement les amorces As1-F/Ts2-R ont amplifié des fragments de tailles proches de celles attendues dans toutes les souches testées.

Certains des fragments (surligné du tableau 25) issus de ces tests ont alors été clonés puis séquencés en vue de déterminer leur identité (figure 33).

A.

Ts2-R

ATGCCTTTGTTTGAATGTCCGCCCGTTTCAAAGAAATTGTCAAAGATCCCCGCCTTCTTAAC
GTTTCAGCACGTCTTGCCAAATGGCTGCGATTGCTTTTTTCAGTCTCATTTCGCGGCGCTGTAT
AGGCAGTGCCGCTGACAGCCTTCACCTCTAGCGCAAACAGCTTGCGGCGGTGCGACTTTGCCG
CTTGGCGTAAGCGGAATTTCTGTCACTTCTGTAAAGTAAGCCGGCACCATATAGTCAGGAAG
CTGTCTGGCCGCATGCTCCCTCAAGCCGGCTGGCGCAAATTCGCTTCCAGGCTTCTGCACAA
CATATGCGCAAACCTCAAGATTTCCCTGTTTCGTTCAGGGCGTGCCAAAACAACGGCTTTTGCA
ACGTCTGGCGCCTGCTGAATCACGGCCTCAATTTCCCCAAGCTCAATTCGGTATCCGCG

As1-F

B.

Ts2-R

ATGCCTTTTTATGAATGTCCGCCGATCATAAAGAAGTTATCGGTAACGCCTGCTTTGACGCC
GAGTATTCCTTCCCAGATCTCCGCTAATTTCTTTTCACTCTCTGTTTCTGGTGCTGCGTATT
CCGGCTGCACGGCTTCTGATTGCGGTGCAGGCAGTTCTTTTTTGTGATTTTCCCGTTTGGC
GTGAGCGGCAGACTGTCCATTTGAATCAGGTGGGCAGGGACCATATAGTCAGGCAAAGACTG
AGACAGTGTTTTTCTCAGCTCACTCACAACCACTTCACGATCCGCCGTGAAGTACGCGCAAA
TTTCATTTTTCATCTGCTTTGTTCTTGCGGATGATCACCGCAGCCTCTGTACATGCTCAGCC
ATGTTTCAGTTTTGTTTCAATTTGCGCAAAGCTCAATCCGGTAGCCGCG

As1-F

C.

Tm1-R

CCGAGATCAAAGAAATTATCGAAACAGTTGCAAGGCTATATCGCCGCAAATTGGAACAAACG
CTTGACGATCATCGCGAAGCTCTTTTAAAGGTGTTTCCGAATTGGAAGACGCGACTTACGG
GGTCTACCTGGACCACCTGCTGCCTCCTGTATTTGAAAGTCTCGAACAGTCCGGCTTTACAA
CGATTCAAAATGCGGGAAAAGGCGATTTTTTTATCGGAAAGGGCTTGAAGTTTAGACAGTCC
ATGGAAAAATGGGGAGCTGACAACTGCCGCTCACGTGTGTTTTGGGTTGTCATCTCCGATCA
GCAAAAACAGCCGGCCGGGACGCTGCTCTTTGACTTTTATCACTCTCACG

Am1-F

Figure 33. Séquence des fragments PCR (A. souche 6 avec As1-F/Ts2-R, B. souche 21 avec As1-F/Ts2-R, C. Souche 21 avec Am1-F/Tm1-R)

Les analyses Blast N et Blast X de ces séquences dans Genbank (NCBI) ont ainsi été réalisées et présentées dans le tableau 26.

Tableau 26. Résultats de Blast des séquences de fragments obtenus par PCR avec les amorces dégénérées

Souches	Abréviation	Taille amplicons	BlastN et X	Identité (%)
6	As1-F/Ts2-R	431 bp	Surfactine synthétase de <i>B. subtilis</i>	98
	Am1-F/Tm1-R	445 bp	2-deoxy-5-kéto-D-gluconic acid 6-phosphate aldolase de <i>B. subtilis</i> 168	98
11	As1-F/Ts2-R	428 bp	Lichenysine C synthétase de <i>B. licheniformis</i> ATCC 14580	99
	Af2-F/Tf1-R	476 bp	G3P perméase de <i>B. licheniformis</i> ATCC 14580	99*
12	As1-F/Ts2-R	419 bp	Surfactine synthétase de <i>B. subtilis</i> 168	96
		425 bp		97
13	As1-F/Ts2-R	419 bp	Surfactine synthétase de <i>B. subtilis</i> 168	98
	Af2-F/Tf1-R	476 bp	G3P perméase de <i>B. licheniformis</i> ATCC 14580	99*
16	As1-F/Ts2-R	419 bp	Surfactine synthétase de <i>Bacillus subtilis</i> 168	96
	Am1-F/Tm1-R	403 bp	Protéine hypothétique de <i>Bacillus licheniformis</i> ATCC 14580	99
21	As1-F/Ts2-R	419 bp	Surfactine synthétase de <i>B. subtilis</i> 168	97
	Am1-F/Tm1-R	445 bp	2-deoxy-5-kéto-D-gluconic acid 6-phosphate aldolase de <i>B. subtilis</i> 168	98

* Pourcentage d'identité ayant des E. value $\neq$ 0.00

Comme attendu, les amorces As1-F/Ts2-R ont mis en évidence la présence de gènes de la surfactine synthétase chez les souches 6, 12, 13, 16 et 21 avec une homologie avoisinant 98%. Elles détectent aussi la présence des gènes de la lichenysine synthétase chez la souche 11 avec une homologie de 99%.

Par contre, les amorces Am1-F/Tm1-F et Af2-F/Tf1-R n'ont pas permis de détecter des gènes NRPS comme espéré. Elles ont par contre amplifié des fragments identifiés comme présents dans des souches de *Bacillus* et qui seraient impliqués dans le métabolisme des carbohydrates.

Conclusions sur le criblage moléculaire des lipopeptides

Ces amorces montrent la présence de gènes impliqués dans la synthèse de lipopeptides (surfactine ou lichenysine) dans les souches d'intérêt (6, 11, 12, 13, 16 et 21).

Malgré l'amplification de fragments de taille proche des tailles attendues par les amorces spécifiques Am1-F/Tm1-F et Af2-F/Tf1-R, aucun gène impliqué dans la synthèse de fengycine et d'iturine n'a pu être mis en évidence. En effet, lors d'une PCR, lorsque le gène recherché est absent du génome bactérien exploré, les amorces de nature dégénérée, hybrident d'autres fragments non spécifiques qui pourraient avoir une taille proche ou totalement différente de celle

attendue. Le séquençage alors des amplicons est alors le moyen ultime d'identifier la nature de ce fragment.

Conclusions et Perspectives

La combinaison des méthodes analytiques classiques (activités biologiques et recherche de molécules spécifiques) et des méthodes moléculaires de criblage par PCR a permis la mise en évidence de bacilles producteurs de surfactines dans les aliments exotiques: Soumbala et Bikalga. Le croisement de ces analyses confirme la présence de ces lipopeptides dans les souches 6 et 12. En effet, les analyses biochimiques par Maldi-TOF n'ont pas permis de détecter des lipopeptides dans les souches 11, 13, 16 et 21 alors que le criblage moléculaire a permis leur mise en évidence, et cela en conformité avec les tests d'activités biologiques et la mesure des tensions de surface. Le pouvoir antifongique exprimé par la plupart de ces souches n'est alors pas le fait de molécules iturines ou fengycines puisqu'aucune des techniques n'a permis leur détection. Celui-ci pourrait être dû à la surfactine qui a été aussi décrite comme possédant un faible pouvoir antifongique (Ullrich *et al.* 1991).

Ces travaux montrent pour la première fois la présence de souches productrices de lipopeptides de type non ribosomal dans ces aliments.

Ces molécules si elles sont présentes dans le Soumbala et le Bikalga pourraient en outre être impliquées dans la bioconservation de ces aliments.

Valorisation

✓ **Aly Savadogo, Arthur Tapi, Marlène Chollet-Imbert, Bernard Wathelet and Philippe Jacques** (2009) Detection of surfactins producing strains in Soumbala and Bikalga fermented condiments using Polymerase Chain Reaction and Matrix Assisted Laser Desorption/Ionization-Mass Spectrometry methods (Soumis dans International Journal of Food Microbiology)

Chapitre 8 Mise au point du logiciel PrimerDeg 0.1

Dans le but d'accélérer la procédure fastidieuse et consommatrice de temps dans la définition des amorces dégénérées présentée dans le Chapitre 6, un outil bioinformatique a été développé et nommé PrimerDeg 0.1 (numéro 1 de la version 0).

Le logiciel PrimerDeg 0.1 comporte deux parties essentielles qui permettent l'exécution des deux actions principales mises en œuvre dans la définition des amorces dégénérées. La première partie du logiciel, nommée ProtRegion2DNA est consacrée à l'extraction et l'alignement des domaines nucléiques à étudier. Le second volet appelé PrimerDeg fournit à l'utilisateur, la région la plus appropriée qui pourra servir dans la définition des amorces.

Il s'agit à terme de mettre au point soit sous forme d'un logiciel téléchargeable, soit sous forme d'une application web, une interface conviviale qui permette de réaliser les 4 différentes étapes critiques présentées dans la figure 34 pour la définition des amorces dégénérées spécifiques.

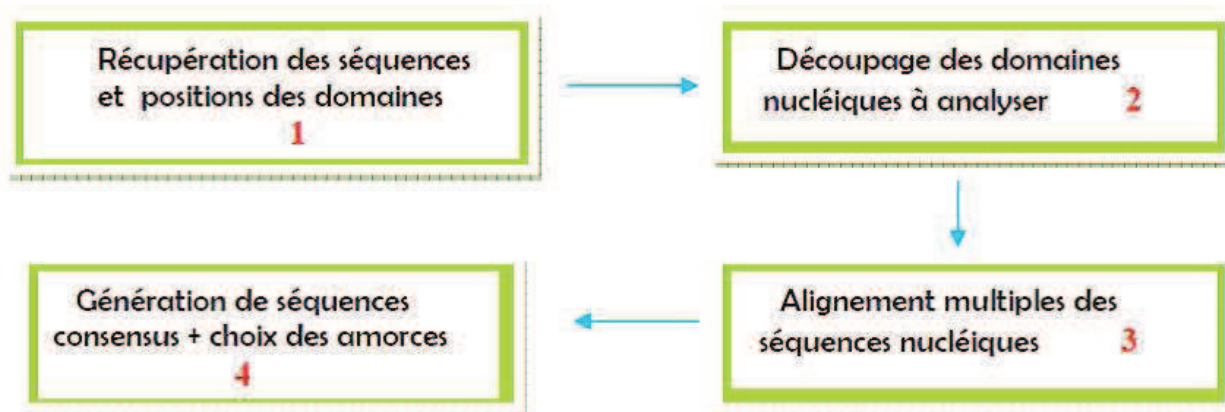


Figure 34. Protocole général de fonctionnement du logiciel PrimerDeg 0.1 (en chiffre rouge, les différentes étapes critiques mises en œuvre pour définir des amorces)

Dans ce chapitre, nous présentons la conception de l'outil, son état d'avancement par rapport aux points critiques mentionnés dans la figure 34, son utilisation et l'interprétation des résultats qu'il génère. Le but final étant que le logiciel regroupe sur une seule interface, l'ensemble de ces procédures pour définir des amorces dégénérées capables de détecter des gènes en particulier de type NRPS.

I. Protocole général du fonctionnement de PrimerDeg 0.1

La récupération de la séquence nucléique correspondant à la protéine non ribosomiale et la détermination des positions des séquences nucléiques correspondant aux domaines NRPS à utiliser constituent la première étape dans la définition des amorces.

1.1. Récupération des séquences

Il est possible, connaissant le numéro d'identification (numéro d'accèsion) ou le nom littéral d'une protéine (exemple : surfactine synthetase C) d'avoir accès à sa séquence protéique et nucléique dans la banque Genbank de NCBI (Benson *et al.* 2008).

L'utilisateur doit accéder à ces informations qui sont nécessaires pour l'utilisation du logiciel. En effet, NCBI référence des domaines NRPS lorsqu'ils sont présents dans une protéine sous forme de régions et mentionne les différentes positions de début et de fin de celles-ci. Il référence aussi d'autres domaines non NRPS. L'analyse des fonctions afférentes et la nature de chacune des régions a permis d'établir une corrélation entre les domaines NRPS et les régions proposées par NCBI. Ainsi, la région condensation correspond au domaine de condensation, la région LuxE au domaine d'Adénylation, la PP-binding au domaine de Thiolation. L'utilisateur choisit selon le découpage proposé, les positions des domaines NRPS qu'il souhaite utiliser ou peut même choisir d'autres découpages provenant d'autres outils comme le logiciel NRPS-ProBioGEM (non publié). L'accès à la séquence nucléique correspondant à la protéine découpée dont les régions vont servir à définir des amorces se fait en cliquant sur le bouton CDS au bas de la séquence protéique.

Ce travail de récupération des positions des domaines spécifiques et de la sélection des séquences n'est pas effectué par PrimerDeg 0.1.

1.2. Détermination de la position et découpage des séquences

Connaissant les positions de début et de fin de chaque région sur la séquence protéique, il est possible de déterminer sa séquence nucléique correspondante sur la CDS.

Lors de la définition des amorces dégénérées dans les chapitres précédents (absence du logiciel PrimerDeg), l'obtention de la séquence nucléotidique correspondante s'est faite en multipliant par 3 (chiffre correspondant au nombre d'acides nucléiques codant un acide aminé selon le code génétique) les différentes positions d'une région protéique donnée par simple translation des valeurs obtenues sur la CDS.

La figure 35A montre un exemple des positions des domaines A et T sur la protéine de la synthétase de surfactine A et leur correspondant en séquences nucléique sur la CDS (Figure 35B).

A. Séquence protéique et position de début et de fin des domaines A et T

gi|225184732|emb|CAB12142.2| surfactine synthétase [B. subtilis str. 168]

301 NIDGGQAFSE FVKDRMKDLM KTLRHQKYPY NLLINDLRET KSSLTKLFTV SLEYQVMQWQ
361 KEEDLAFLTE PIFSGSGLND VSIHVKDRWD TGKLTIDFDY RTDLFSREEI NMICERMITM

Domaine d'adénylation (478-956)

421 LENALTHPEH TIDELTLISD AEKEKLLARA GGKSVSYRKD MTIPELFQEK AELLSDHPAV
481 VFEDRTLSTYR TLHEQSARIA NVLKQKGVGP DSPVAVLIER SERMITAIMG ILKAGGAYVP
541 IDPGFPAERI QYILEDGAD FILTESKVAA PEADAELIDL DQAIIEGAEI SLNADVARN
601 LAYIIYTSYT TGRPKGVMIE HRQVHLLVES LQQTIIYQSGS QTLRMALLAP FHFDAVSKQI
661 FASLLLGQTL YIVPKKTVTN GAALTAYYRK NSIEATDGRP AHLQMLAAAG DFEGKLKHKM
721 LIGGEGLSV VADKLLKLFK EAGTAPRLTN VYGPTETCVD ASVHPVIPEN AVQSAYVPIG
781 KALGNNRLYI LDQKGRLOPE GVAGELYIAG DGVGRGYLHL PELTEEKFLQ DPFVPGDRMY
841 RTGDVVRWLP DGTIEYLGRE DDQVKVRGYR IELGEIEAVI QQAPDVAKAV VLARPDEQGN
901 LEVCAYVVQK PGSEFAPAGL REHAARQLPD YMPAYFTEV TEIPLTPSGK VDRRKLFALE

Domaine de Thiolation (978-1040)

961 VKAVSGTAYT APRNETEKAI AAIWQDVLNV EKAGIFDNFF ETGGHSLKAM TLLTKIHKET
1021 GIEIPLQFLF EHPTITALAE EADHRESKAF AVIEPAEKQE HYPLSLAQQR TYIVSQFEDA
1081 GVGYNMPAAA ILEGPLDIQK LERAFQGLIR RHESLRTSFV LENSTPRQKI HDSVDFNIEM
1141 IERGGRSDEA IMASFVRTFD LAKAPLFRIG LLGLEENRHM LLFDMHHLIS DGVSIGIMLE
1201 ELARIYKGEQ LPDLRLQYKD YAVWQSRQAA EGYKKDQAYW KEVFAGELPV LQLLSDYPRP

B.

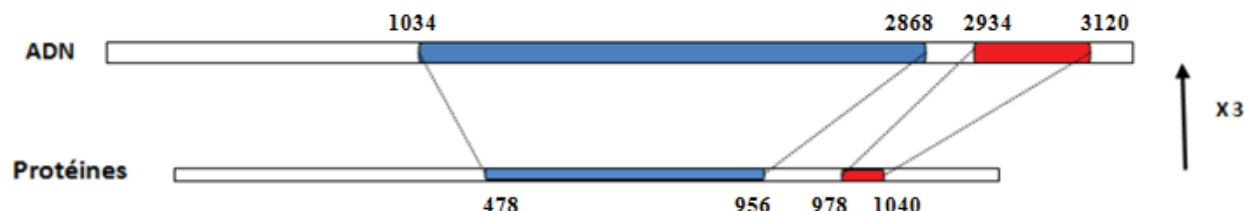


Figure 35. A. Un extrait de la séquence protéique de la surfactine synthétase A (en bleu : le domaine d'adénylation débute à la position 478 et prend fin à l'acide aminé en position 956, en rouge : le domaine de thiolation débute à la position 978 et prend fin à l'acide aminé en position 1040)

B. Représentation schématique et positionnement des domaines d'adénylation et de thiolation sur la séquence protéique et la séquence nucléotidique correspondante

Une formule de Conversion Des Positions (FCDP) a été déterminée pour calculer le début et la fin de la séquence nucléique correspondante au domaine protéique:

$$\begin{aligned} \text{FCDP (Début de la séquence nucléique)} &= \text{début de la CDS} + (\text{début de région protéique} - 1) \times 3 \\ \text{FCDP (Fin de la séquence nucléique)} &= \text{début de la CDS} + \text{fin de région protéique} \times 3 - 1 \end{aligned}$$

Le logiciel dans sa conception utilisera alors cette formule pour extraire la séquence nucléique d'intérêt lorsque l'utilisateur lui aura donné les différentes positions au niveau protéique des domaines à utiliser. Cette formule permet au logiciel de gérer une séquence pour laquelle la CDS serait inclus dans une séquence nucléique plus grande.

Le découpage de la CDS pour l'obtention des séquences nucléiques qui serviront pour les alignements (étape 2 du protocole général de la figure 34) est entièrement automatisé.

1.3. Alignements multiples des séquences nucléiques

Après extraction des séquences nucléiques correspondant aux domaines protéiques, la 3^{ème} étape consiste à les aligner.

Des logiciels webs et libres (Clustal, Mafft, etc...) permettent de réaliser des alignements multiples de séquences nucléiques dans le but de déterminer des régions conservées. Les alignements de séquences sont des comparaisons entre des séquences primaires (ADN ou protéines) pour identifier les zones de concordance qui traduisent des similarités ou des dissemblances.

Compte tenu de la multiplicité des logiciels existants déjà dédiés aux alignements de séquences, cette 3^{ème} étape sera réalisée par un logiciel extérieur appelé Jalview (Waterhouse *et al.* 2009) par une applet permettant son accès directement sur PrimerDeg 0.1. Le logiciel Jalview est un éditeur qui permet à la fois de lancer directement des logiciels d'alignements, de visualiser et de corriger manuellement ces alignements.

Le motif conservé ainsi obtenu sera utilisé pour définir une séquence consensus différente de celle générée par Jalview selon les calculs ci-après.

1.4. Calcul du consensus

La séquence consensus est une représentation de l'alignement sur la base des règles établies par Cavener (1987).

Nous présentons dans ce paragraphe, le principe du calcul du logiciel pour exécuter l'étape 4 du protocole général de la figure 34 pour générer une séquence consensus.

Pour exemple, un alignement est réalisé avec 15 séquences différentes de 10 nucléotides chacune et présenté à la figure 36.

	1	2	3	4	5	6	7	8	9	10
Seq ₁	G	C	C	G	G	A	A	G	T	G
Seq ₂	A	C	C	G	G	A	A	G	C	A
Seq ₃	G	C	C	G	G	A	T	G	T	A
Seq ₄	A	C	C	G	G	A	A	G	C	T
Seq ₅	A	C	C	G	G	A	T	A	T	A
Seq ₆	C	C	C	G	G	A	A	G	T	G
Seq ₇	A	C	A	G	G	A	A	G	T	C
Seq ₈	G	C	C	G	G	A	T	G	C	A
Seq ₉	T	C	C	G	G	A	A	G	T	A
Seq ₁₀	A	C	A	G	G	A	A	G	C	G
Seq ₁₁	A	C	A	G	G	A	T	A	T	G
Seq ₁₂	T	C	C	G	G	A	A	A	C	C
Seq ₁₃	A	C	A	G	G	A	T	A	G	C
Seq ₁₄	C	A	A	G	G	A	C	G	A	C
Seq ₁₅	T	C	T	G	G	A	C	C	C	T

Figure 36. Alignement multiple de séquences nucléiques

La séquence consensus est construite par l'analyse des colonnes de l'alignement multiple. Les séquences de l'alignement sont conservées au sein d'une colonne si celle-ci contient une proportion importante de nucléotides identiques.

Dans notre exemple de la figure 36, le nombre total de nucléotides par colonne (N) est égal à 15. L'ensemble de toutes les colonnes forme un motif.

Le Nombre de chaque Nucléotide par Colonne (NNC) est le nombre de fois où un nucléotide est présent dans une colonne. Le tableau 27 présente les NNC de notre exemple d'alignement de la figure 36.

Tableau 27. Répartition du nombre de nucléotides par colonne

		Colonnes									
		1	2	3	4	5	6	7	8	9	10
Nucléotides	A	7	1	5	0	0	15	8	4	1	5
	C	2	14	9	0	0	0	2	1	6	4
	G	3	0	0	15	15	0	0	10	1	4
	T	3	0	1	0	0	0	5	0	7	2

La séquence consensus sera déterminée par la fréquence selon laquelle un nucléotide donné apparait dans une colonne. On parle de **Fréquence De Nucléotides (FDN)** ou Position Frequency Matrix (PFM) qui est le rapport entre le NNC et le nombre total de nucléotides par colonne (N). Le tableau 28 présente la détermination des FDN dans le cas de notre exemple de la figure 36.

Tableau 28. Exemple de la détermination des FDN

		Colonnes										
		1	2	3	4	5	6	7	8	9	10	
Fréquence	Nucléotides	A	0,47	0,07	0,33	0	0	1	0,53	0,27	0,07	0,33
		C	0,13	0,93	0,6	0	0	0	0,13	0,07	0,4	0,27
		G	0,2	0	0	1	1	0	0	0,67	0,07	0,27
		T	0,2	0	0,07	0	0	0	0,33	0	0,47	0,13
Séquence consensus		N	C	H	G	G	A	H	G	Y	N	

Le calcul de la FDN est exécuté par le logiciel et les nucléotides sont affichés pour former la séquence consensus selon les critères ci-après:

a. Un seul nucléotide (A, C, G ou T) est affiché si sa fréquence est supérieure à 50% et au moins deux fois plus élevée que celle du deuxième nucléotide le plus fréquent.

C'est le cas des colonnes 2 et 8 du tableau 28 (encadrées en bleu), le logiciel affichera respectivement les nucléotides C et G car ayant des FDN égale à 0,93 et 0,67 dans la séquence consensus.

Par contre, dans la colonne 7 par exemple, le nucléotide 'A' a une FDN supérieure à 50% mais pas $2 \times 0,33$ (double de la FDN plus proche), la règle « a » ne s'applique pas mais plutôt la règle « c » décrite par la suite.

b. Une double-dégénérescence (R, Y, K, M, S ou W) indiquera que deux nucléotides ont une fréquence cumulée supérieure ou égale à 75% et chacun d'eux à une FDN inférieure à 50%.

C'est le cas de la colonne 9 du tableau 28 (encadrée en rouge), le logiciel utilisera les nucléotides C et T puis affichera le nucléotide dégénéré Y correspondant, selon le code IUPAC-DNA

c. Une triple-dégénérescence (B, D, H ou V) est affichée si un seul des nucléotides (A, C, G, T) n'est pas présent dans la séquence

C'est le cas de la colonne 3 (encadrée en vert dans le tableau 28), le logiciel utilisera les nucléotides A, C et T et affichera H dans la séquence consensus selon le code IUPAC-DNA

d. Toutes les autres distributions de fréquence sont représentées par la lettre "N" lorsqu'elles ne correspondent à aucune des trois premières règles comme le cas de la colonne 10 encadrée en orange du tableau 28.

La séquence consensus qui sera générée par le logiciel sur base de ces critères précédents dans le cas de notre exemple est présentée dans le tableau 28.

Cette séquence consensus sera considérée comme la séquence consensus brute par le logiciel. D'autres critères (calcul du HyC, niveau de dégénérescence de l'extrémité 3') en complément de ceux définissant les nucléotides dégénérés ou non, seront prises en compte par le logiciel pour choisir la séquence consensus qui servirait d'amorce.

II. Présentation et utilisation du logiciel

2.1. Présentation de PrimerDeg 0.1

Le logiciel PrimerDeg 0.1 est une interface constituée de deux onglets correspondant aux deux fonctions principales: les étapes 2 et 4 du protocole général de la figure 34. La figure 37 présente la page d'accueil de l'outil.

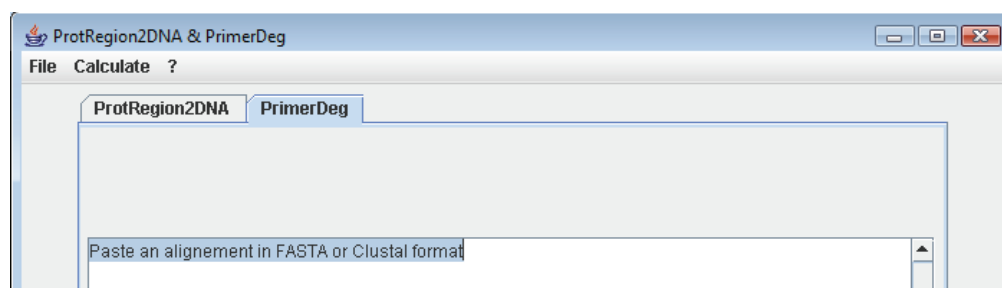


Figure 37. Interface montrant les deux onglets du logiciel PrimerDeg 0.1.

Le premier onglet est ProtRegion2DNA qui permet d'exécuter la deuxième étape, c'est-à-dire l'extraction des différentes séquences nucléiques correspondant aux domaines protéiques. Accessoirement, il permet d'utiliser le logiciel Web Jalview (étape 3) soit directement via internet, soit à l'aide de sa version téléchargeable préalablement installée pour effectuer les alignements multiples.

Le second onglet appelé PrimerDeg génère des séquences consensus qui peuvent être utilisées comme amorces, donc de l'étape 4 du protocole général.

L'interface de PrimerDeg 0.1 présente aussi trois autres menus :

- File : permet de lancer le logiciel d'alignement Jalview ou de quitter le logiciel
- Calculate : permet d'extraire les séquences nucléiques et les alignements réalisés
- ? : renseigne sur l'outil en général (aide, mises à jour, auteurs)

2.2. Utilisation du logiciel

2.2.1. Utilisation de l'onglet ProtRegion2DNA

Les positions des domaines d'intérêt sont copiées par l'utilisateur à partir des données de la banque Genbank (surlignée en jaune, ici les positions de début et de fin d'un domaine A de la synthétase SrfA-A : CAB12142.1, figure 38) puis collées dans la zone de saisie 3 de ProtRegion2DNA (Figure 39). Pour obtenir la séquence nucléotidique du CDS dans NCBI, il faut cliquer sur le lien hypertexte (surligné en vert dans la figure 38).

La séquence et les positions du CDS au niveau nucléique sont respectivement collées dans les zones de saisie 1 et 2 en les séparant par une virgule comme présenté dans la figure 39.

<u>Region</u>	478..956 /region_name="LuxE" /note="Acyl-protein synthetase, LuxE; c110450" /db_xref="CDD:142560"
<u>Region</u>	978..1040 /region_name="PP-binding" /note="Phosphopantetheine attachment site; c109936" /db_xref="CDD:142335"
<u>Region</u>	1058..1346 /region_name="Condensation" /note="Condensation domain; c109290" /db_xref="CDD:142200"
<u>Region</u>	 3429..3564 /region_name="NRPS-para261" /note="non-ribosomal peptide synthase domain TIGR01720" /db_xref="CDD:130781"
<u>CDS</u>	1..3587 /gene="srfAA" /locus_tag="BSU03480"

Figure 38. Exemple de découpage en domaines par NCBI (protéine SrfA-A)

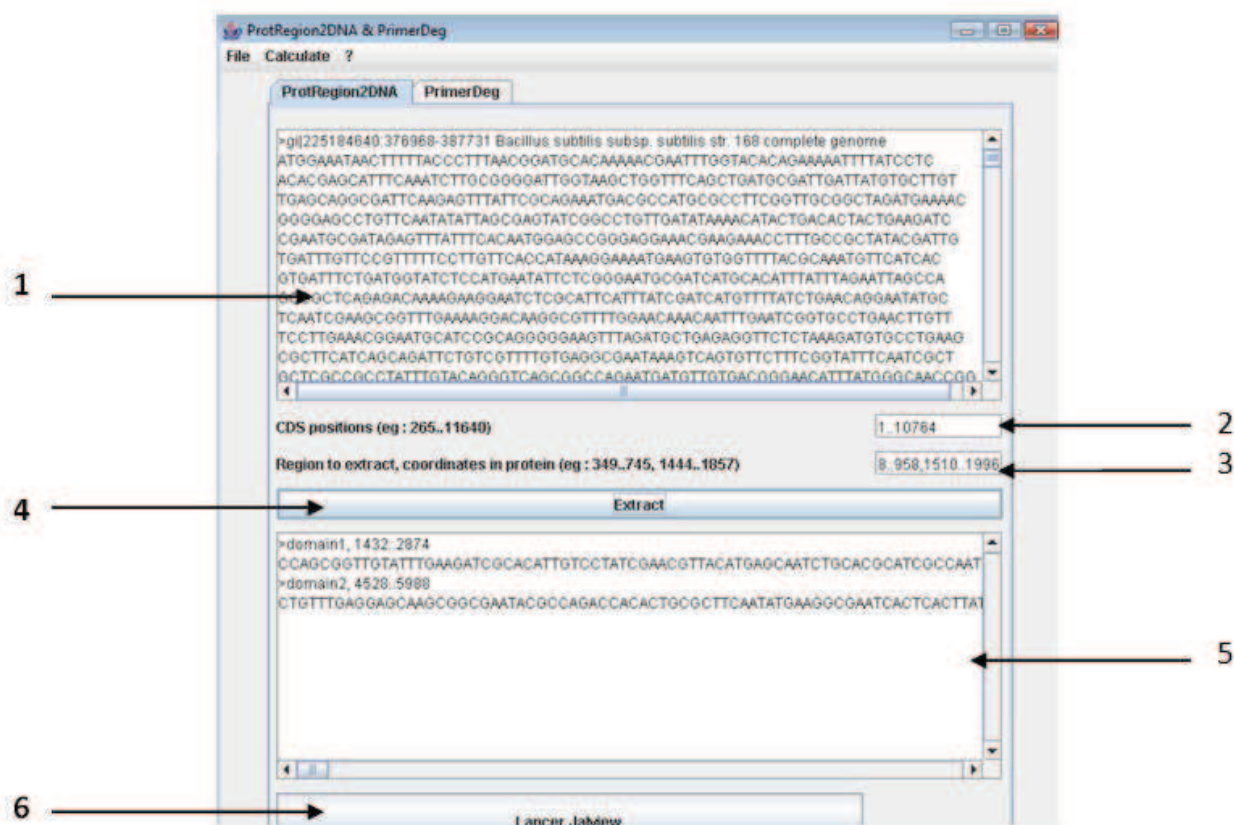


Figure 39. Interface de l'onglet ProtRegion2DNA du logiciel PrimerDeg 0.1

Ensuite en cliquant sur le bouton 4 'Extract', ProtRegion2DNA va extraire les domaines souhaités et les afficher dans la zone de saisie rectangle 5.

Le logiciel trouve alors la séquence nucléique de tous les domaines en exécutant le calcul nécessaire effectué pour afficher la séquence nucléique pour chacun des domaines. Ainsi, cette étape fastidieuse et source d'erreurs n'est plus effectuée manuellement.

Le bouton 6 permet de lancer Jalview de façon automatique. Elle permet de réaliser les alignements des domaines extraits.

Les alignements des séquences nucléiques ainsi obtenus vont servir de matrice pour la définition des séquences consensus par l'onglet PrimerDeg.

2.2.2. Utilisation de l'onglet PrimerDeg

Les alignements de séquences nucléiques obtenus à l'aide de Jalview sont copiés dans le format Fasta ou Clustal puis collés dans la zone de saisie 1 de PrimerDeg (Figure 40). La première étape consiste à déterminer la taille de l'amorce souhaitée dans l'onglet 2 'Window size' (la taille par défaut est de 20 bp).

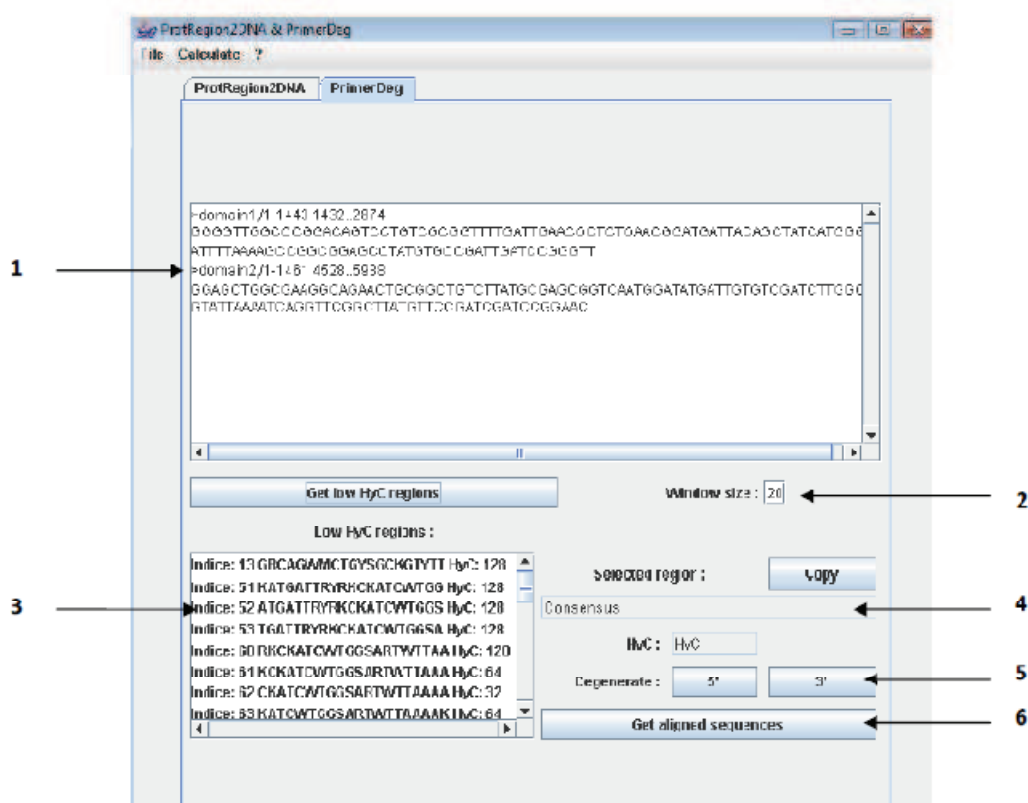


Figure 40. Interface de l'onglet PrimerDeg du logiciel PrimerDeg 0.1

Les séquences consensus sont générées par un clic sur le bouton d'application 'Get Low Hyc regions'. En effet, le logiciel détermine les séquences consensus en tenant compte à la fois des critères de sélection des nucléotides dégénérés (§ 1.1.4 de ce Chapitre) et du Coefficient d'Hybridation (HyC) en utilisant le motif le plus conservé. Le HyC est déterminé selon la formule suivante décrite dans le § 1.1.2 du Chapitre 5.

$$HyC = 1^{ni} \times 2^{(nR+nY+nM+nK+nW+nS)} \times 3^{(nB+nD+nH+nV)}$$

Le HyC de chaque séquence consensus dite brute doit être inférieur à 200.

Plusieurs séquences consensus sont ainsi affichées dans l'espace 3 'Start Consensus HyC' et représentent chacune une proposition d'amorce. Le choix de la séquence consensus qui servira d'amorce est basé sur le HyC le plus faible et une extrémité 3' la moins dégénérée (possède le moins de nucléotides dégénérés).

La séquence consensus est affichée dans la barre consensus 4 en cliquant dessus. Elle peut être copiée et analysée pour vérifier son efficacité.

Parfois, le HyC de la séquence la plus consensuelle est supérieur à 200, les boutons 5' et 3' recalculent les FDN en éliminant en priorité les nucléotides de l'extrémité 3' de l'amorce qui permettront au logiciel d'utiliser un autre critère. Par exemple, lorsque le nucléotide présent dans une séquence consensus l'a été avec l'utilisation de la règle « c », le logiciel supprimera un nucléotide de la colonne d'alignement pour utiliser la règle précédente (soit dans ce cas la règle « b » soit la règle « a » du § 1.1.4 de ce Chapitre) pour afficher un autre nucléotide. Cela diminue *de facto* le HyC de la séquence consensus.

Conclusion et Perspectives

A l'opposé de plusieurs logiciels connus (Primer3', Oligo 6, PerlPrimer, etc...) permettant la définition d'amorces non dégénérées pour la PCR, tout comme CODEHOP (Rose *et al.* 2003) qui est spécialisé dans la définition des amorces dégénérées à partir des alignements de séquences d'acides aminés conservés, PrimerDeg 0.1 est un outil aidant à la définition des amorces dégénérées à partir des séquences nucléiques, en particulier des synthétases non ribosomiales.

La version zéro (0) de PrimerDeg 0.1 bien qu'étant une application fonctionnelle reste encore à améliorer à la fois dans sa conception et dans ses fonctionnalités. En effet, le type de découpage des domaines utilisés par le logiciel à partir de l'extraction sur NCBI pourrait constituer une faiblesse car ne permettant pas l'affichage de certains domaines dits secondaires.

D'autres fonctionnalités doivent être rajoutées pour atteindre les objectifs fixés notamment pour assurer l'autonomie de cet outil c'est-à-dire sans avoir recours à d'autres plateformes comme JalView. Il devra aussi à terme permettre le stockage et la validation des amorces.

Valorisations

- ✓ **Tapi A, Pupin M, Chollet M, Jacques Ph** PrimerDeg 0.1: Software to define degenerated primers for PCR. Mis à disposition du logiciel sur la plate forme bioinformatique du Laboratoire ProBioGEM.

Chapitre 9 Identification d'un nouveau lipopeptide non ribosomal chez *Bacillus thuringiensis* serovar *berliner* : la kurstakine

Un fragment de gène codant une synthétase a été détecté lors des chapitres précédents par les amorces NRPS-lipopeptides dans le génome de la souche de *Bacillus thuringiensis* serovar *berliner* 1915 (Institut Pasteur de Paris, CRBIP, France).

Dans ce chapitre, nous combinons à la fois les méthodes moléculaires que nous avons mises au point pour la détection de molécules non ribosomiales, et des techniques d'identification par HPLC et spectrométrie de masse, dans le but de déterminer la nature de la molécule issue de cette synthétase. Le potentiel de synthèse de molécules non ribosomiales par cette souche n'est en effet pas connu. A cette étude, nous ajoutons une autre souche, *Bacillus thuringiensis* S 45, isolée de céréales et offerte par l'Université d'Alger (Algérie), car décrite comme possédant des activités biologiques notamment antifongique. L'objectif est de vérifier si le gène détecté dans la souche *berliner* 1915 est également présent dans la souche S 45 et si celui-ci coderait pour les molécules responsables des activités biologiques.

I. Analyses biologiques et biochimiques

Les activités antifongiques ont été confirmées pour les deux espèces de *B. thuringiensis* (BT S 45 et BT 1915), comme décrit dans le § 2.3.3 du Chapitre Matériels et Méthodes. La figure 41 présente une photographie des tests antifongiques réalisés.

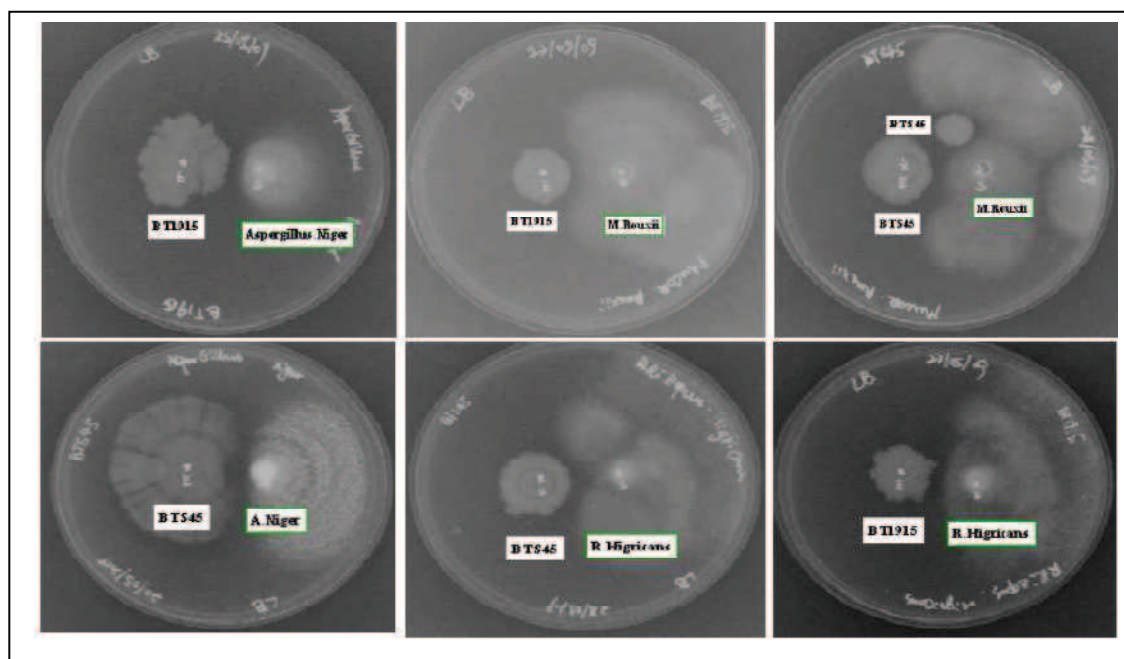


Figure 41.
Photographie
des tests
antifongiques
BT S 45 et
BT 1915

Les résultats des tests antifongiques obtenus pour les deux souches sont résumés dans le tableau 29.

Tableau 29. Activités antifongiques des souches de *Bacillus thuringiensis*

	Souches cibles	<i>M. rouxii</i>	<i>P. roquefortii</i>	<i>R. nigricans</i>	<i>A. niger</i>	<i>A. flavus</i>
Souches testées	BT S 45	+	-	+	+	+
	BT 1915	+	-	+	+	+

– absence d'inhibition, + activité inhibitrice de la souche

Les deux souches ont exprimé une activité antifongique similaire contre les moisissures testées. La détection de fragment par des amorces spécifiques définies à partir de gènes NRPS de lipopeptides et, les activités antifongiques indiquent que les molécules synthétisées par les souches de *B. thuringiensis* pourraient être de type lipopeptidique.

1.1. Recherche de lipopeptides

Des analyses biochimiques (détection après séparation par chromatographie en phase inverse ou par spectrométrie de masse) ont été réalisées pour caractériser dans les surnageants de culture ou directement sur colonie, la présence de lipopeptides.

1.1.1. A partir de surnageant de culture

- Par HPLC

Des surnageants de culture des deux souches dans le milieu DSM ont été purifiés sur colonnes Altech C18 comme décrit dans le § 2.4.2 du Chapitre Matériels et Méthodes.

Les échantillons purifiés ont ensuite été analysés par chromatographie liquide à haute performance sur phase inverse dans des conditions d'analyse permettant de détecter les trois familles de lipopeptides produits par *Bacillus subtilis*.

Les chromatogrames obtenus montrent qu'aucun pic n'a été détecté lors de l'utilisation des gradients spécifiques de surfactines et de fengycines pour les deux souches.

La figure 42 présente les pics obtenus avec le gradient iturine pour les deux souches comparé à un standard d'iturine.

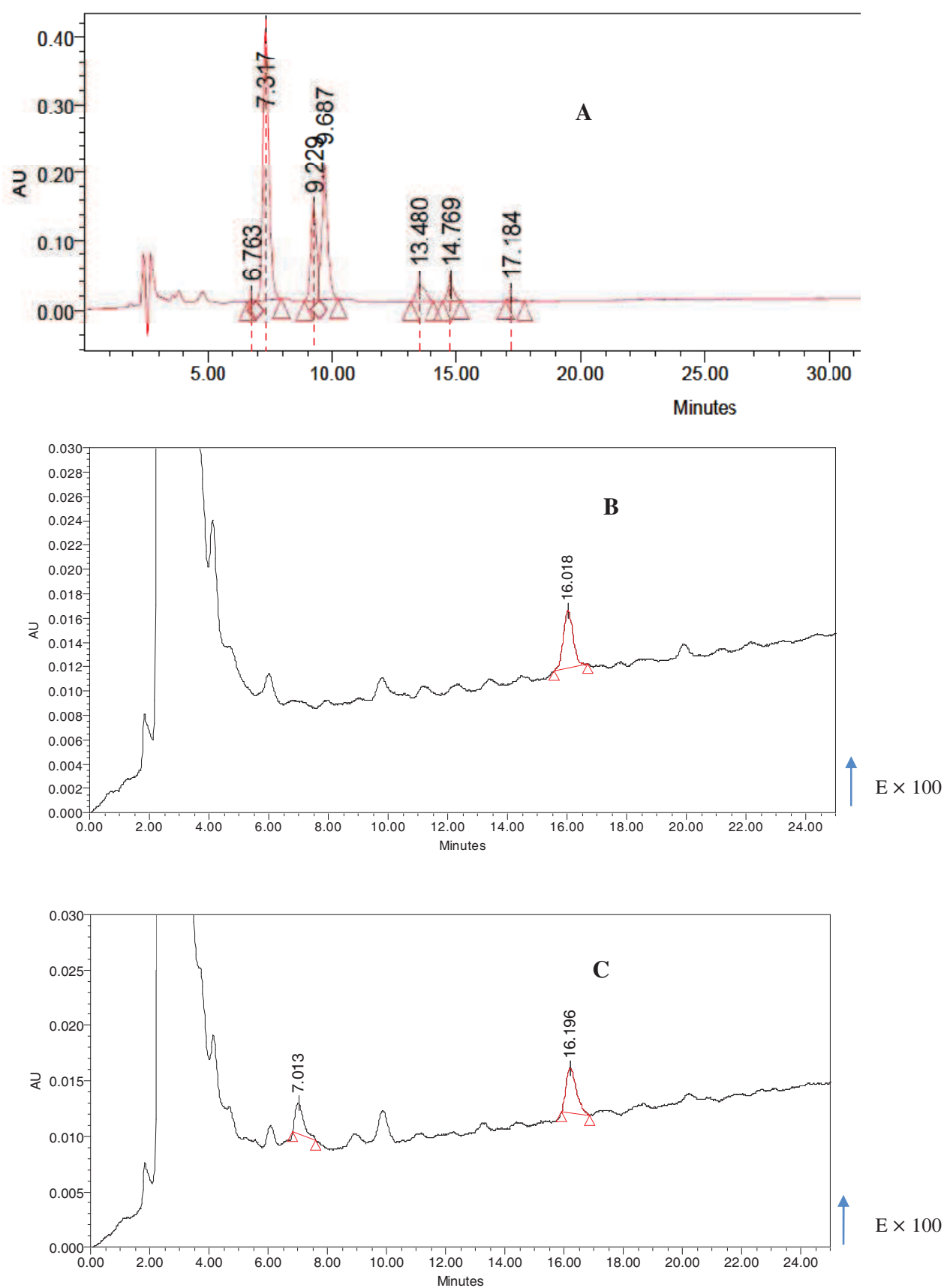


Figure 42. Détection de la présence d'iturines dans des surnageants de culture des souches *B. thuringiensis* serovar *berliner* 1915 et S 45 (**A.** Témoin standard iturine A concentré à 500 mg.mL⁻¹; **B.** souche 1915; **C.** souche S 45)

Les figures 42B et 42C montrent la présence d'un pic (avec un temps de rétention autour de 16 min) dans la zone de rétention des iturines dans les surnageants des deux souches. Ce pic pourrait correspondre à la présence d'une molécule proche de la famille des iturines. Toutefois, la concentration est trop faible pour confirmer au travers d'une analyse spectrale la nature iturinique du composé élué.

- Par Maldi-TOF

Une analyse par spectrométrie de masse Maldi-TOF a été réalisée en collaboration avec le Pr. Bernard Wathelet de Gembloux Agro Bio Tech (Lenders *et al.* 1999) soit directement à partir des colonies ensemencées sur milieu DSM solide soit à partir de surnageants de culture réalisé dans du milieu DSM liquide.

Le tableau 30 et la figure 43 présentent respectivement les molécules correspondant aux masses des pics détectés en milieu solide et quelques spectres qui ont été obtenus à partir des souches testées en milieux solide et liquide.

Tableau 30. Masses des pics détectés sur colonie (C) et dans les surnageants de culture (S) de souches *B. thuringiensis* S 45 et 1915 obtenus par Maldi-TOF

Molécules détectées	Kurstakines		Surfactines		Iturines		Fengycines	
	*S	*C	S	C	S	C	S	C
BT S 45	Abs	C11 C12 C13	Abs	Abs	Abs	C14	Abs	Abs
BT 1915	Abs	Abs	Abs	Abs	Abs	Abs	Abs	Abs

Abs : absence de pics, S : surnageant, C : colonies

Un massif de pics et un pic isolé correspondant à des rapports m/z de lipopeptides connus ont pu être identifiés : m/z 878–944 correspondent à la région des kurstakines et m/z 1065 correspond à l'ion $M+Na^+$ de l'iturine C14.

La présence des molécules de masses similaires à celles des kurstakines a été mise en évidence chez la souche S 45. En effet, les m/z des pics correspondant aux masses de la kurstakine et à ces homologues C11, C12 et C13 ont été détectés dans la souche S 45 en milieu solide. Des masses des pics proches de celles de la kurstakine ont été mises en évidence dans les surnageants des deux souches.

Cette molécule avait déjà été mise en évidence lors d'un criblage de lipopeptides dans le genre *Bacillus* (Hathout *et al.* 2000, Price *et al.* 2007).

La figure 43 présente les spectres obtenus en milieux solide et liquide des deux souches.

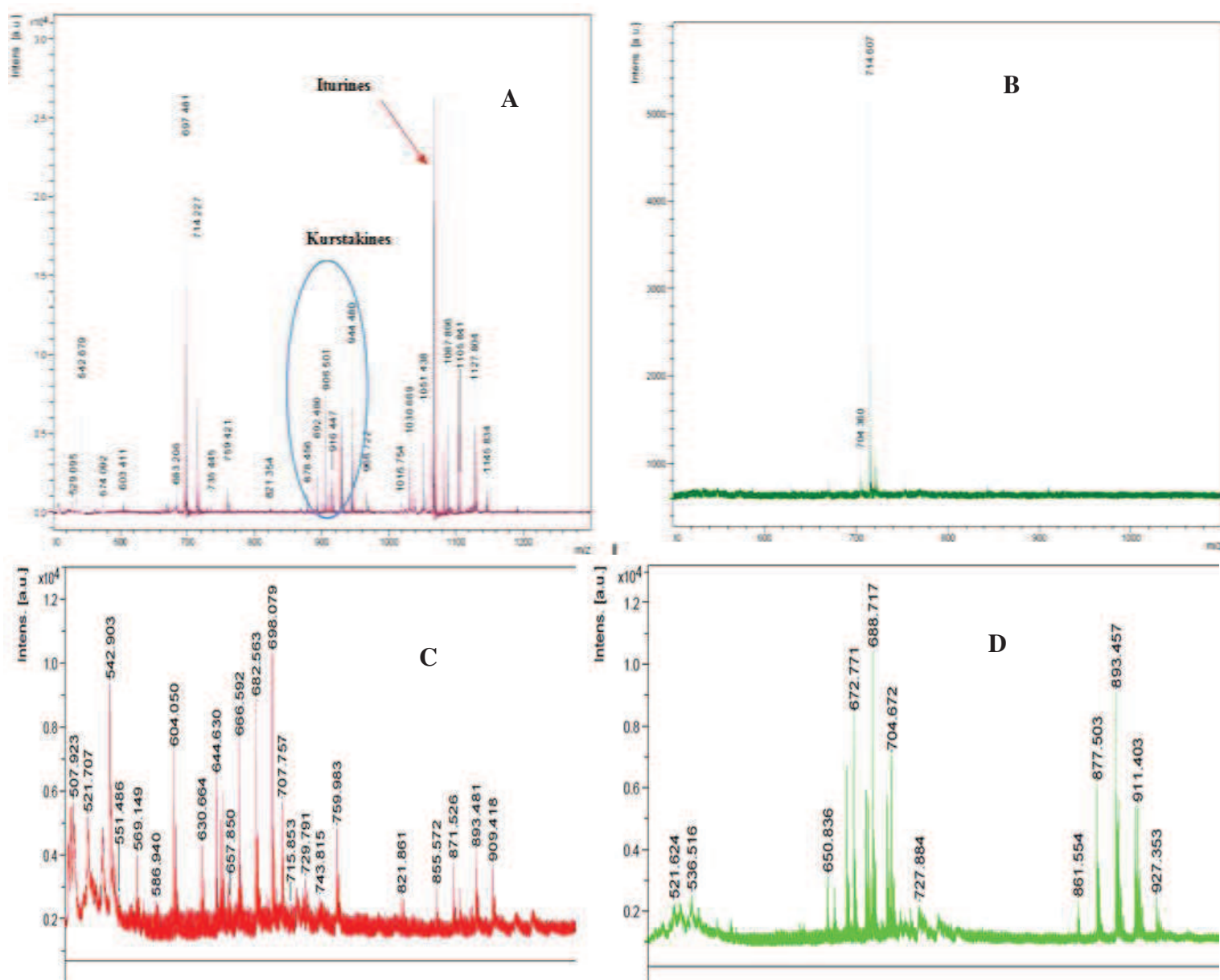


Figure 43. Spectres obtenus sur milieux DSM solide (**A.** *B. thuringiensis* S 45 ; **B.** *B. thuringiensis* 1915) et liquide (**C.** *B. thuringiensis* S 45 ; **D.** *B. thuringiensis* berliner 1915)

Des analyses moléculaires seront donc effectuées pour rechercher les gènes de la molécule de la famille des iturines détectée chez BT S 45 (figure 43 A). Ce travail ne pourra pas être réalisé pour la kurstakine car cette molécule n'est pas connue comme étant de type non ribosomal et de plus les gènes qui sont impliqués dans sa synthèse ne sont pas séquencés. Les analyses pour la confirmation de la kurstakine ont alors été faites au moyen d'outils bioinformatiques.

1.1.2. Analyses moléculaires

- Recherche du gène de la molécule de la famille des iturines

Les analyses par chromatographie et spectrométrie de masse ont montré la présence éventuelle d'une molécule proche des iturines chez *B. thuringiensis* S 45.

Trois lipopeptides (iturines, mycosubtilines et bacillomycines) constituent la famille des iturines. L'objectif est d'utiliser des amorces spécifiques pour détecter le gène de la synthétase des iturines qui aurait été mis en évidence. Outre les amorces Am1-F/Tm1-F (§ 1.1.2 du Chapitre 6) précédemment définies pour détecter les mycosubtilines et les iturines, de nouvelles amorces ont été mises au point sur les gènes codant la bacillomycine D synthétase chez *Bacillus subtilis* AU195 (Moyne *et al.* 2004).

En effet, une analyse comparative a été réalisée au niveau nucléique et protéique des gènes codant des synthétases des iturines, des mycosubtilines et de la bacillomycine D et de leur produit. Cette analyse révèle que les gènes codant pour les iturine synthétases et les mycosubtiline synthétases ont plus de similarité entre eux que par rapport à ceux codant les bacillomycine synthétases (Tableau 31).

Tableau 31. Analyse comparative des synthétases de la famille des lipopeptides ituriniques

Lipoptides NRPS	Similarité protéique (%)	Similarité nucléique (%)
mycA/ituA	88,7	80,7
mycB/ituB	88,6	79,8
mycC/ituC	79,0	69,7
mycA/BmyA	88,9	42,6
mycB/BmyB	82,5	43,0
mycC/BmyC	85,2	43,9

De nouvelles amorces ont donc été déterminées à partir de l'opéron bacillomycine (AAN07012, AAN07013, AAN07014) de *B. subtilis* AU195 comme décrit dans le Chapitre 6. Les caractéristiques de ces amorces sont présentées dans le tableau 32.

Tableau 32. Caractéristiques des amorces spécifiques de bacillomycine déterminées à l'aide du logiciel AmplifX 1.4.4 (conditions de match par défaut)

Nom amorces	Séquences 5' 3'	HyC	Taille des fragments attendus (bp)
Ab11-F	GATSAWCARGTGAAAATYCG	16	428, 431, 434
Tb11-R	ATCGAATSKCCGCCRARATCRAA	32	

Les amorces Am1-F/Tm1-R ont été utilisées pour la recherche des iturines et des mycosubtilines puis le couple Abl1-F/Tbl1-R pour détecter le gène de la bacillomycine synthétase par PCR avec l'ADN des deux souches S 45 et 1915. Ces différentes amorces spécifiques augmenteront nos chances de détecter les gènes de toutes les molécules appartenant à la famille des iturines.

Les conditions expérimentales d'utilisation de ces amorces sont présentées dans le tableau 5.

Les deux couples d'amorce (Am1-F/Tm1-R et Abl1-F/Tbl1-R) ont été testés *in vitro* sur l'ADN des souches de *B. subtilis* 168, ATCC 6633, *B. thuringiensis* S 45 et 1915. Les souches *B. subtilis* ATCC 6633 et *B. subtilis* 168 ont servi respectivement de contrôle positif et négatif.

Le tableau 33 présente les résultats obtenus par PCR.

Tableau 33. Résultats PCR *in vitro* par les amorces Am1-F/Tm1-R et Abl1-F/Tbl1-R (- pas de bandes, + présence de bandes)

Souches	BT S 45	BT 1915	B.S 6633	B.S 168
Am1-F/Tm1-R	-	-	+	-
Abl1-F/Tbl1-R	+	+	+	+

Aucun gène de type iturine ou mycosubtiline n'a été détecté chez les souches de *B. thuringiensis* étudiées. Ces résultats tendent à démontrer l'absence des gènes codant les synthétases de la mycosubtiline et de l'iturine dans les souches *B. thuringiensis* serovar *berliner* 1915 et *B. subtilis* 168, et confirment la présence de mycosubtiline synthétases chez *B. subtilis* ATCC 6633 (Tableau 15).

Les amorces Abl1-F et Tbl1-R ont mis en évidence des fragments de tailles attendues dans toutes les souches testées, cela semble indiquer que ces souches contiennent des gènes de bacillomycine. L'ensemble des fragments ainsi amplifiés par les amorces Abl1-F/Tbl1-R ont été clonés, séquencés puis identifiés par Blast dans la banque Genbank (NCBI). Les séquences obtenues sont présentées à la figure 44.

A.

Tb11-R

ATCGAATGTCCGCCGAGATCGAA^{AAAA}TGAATCTTGAATTCCGATGTTTTCCATACCTAAAC
TTCATGCCATATCGAAACGATTGTTTTTTCCGTAGACGTACGAGGAGCAATATAACCACTTG
CTATCTCCTGTCTCTCTGGTGCTGGTAAGGCCTTCCTATCAATTTTCCCATTAGGAGTGAGT
GGAAATGCATTTAATCCAACAAAATAGGCTGGAATCATATAATTTGGTAGTTGATCCTTGAG
ATAATCACGCCACTTTTGAGCATCCCCGTCTCCTACAACATAGGCGATCAAACGCGGATCGC
CAGGATAATCTTCCTGCACCATCACAAATAACTTCGTTAACCAGTGTATGCCTTTGTAATGTC
GCTTCAATTTCTCCTAGTTCAATCCGGAACCCAC^{GAATTTTCACTTGATGATC}

Abl1-F

B.

Tb11-R

ATCGAATGTCCGCCGAGATCAA^{AAAA}ACGAATCTTGAATTCCGATGTTTTCTTACCTAAAC
TTCATGCCATATCGAAACGATTGTTTTTTCCGTAGACGTACGAGGAGCAATATAACCACTTG
CTATCTCCTGTCTCTCTGGCGCTGGTAAGGCCTTCCTATCAATTTTCCCATTAGGAGTGAGT
GGAAATGCATTTAATCCAACAAAATAGGCTGGAATCATATAATTTGGTAGTTGATCCTTGAG
ATAATCACGCCACTTTTGAGCATCCCCGTCTCCTACAACATAGGCAATTAAACGTGGGTGCG
CAGGATAATCTTCCCGCACCATCACAAATAGCTTCGTTGACTAGTGTATGCCTTTGTAATGTC
GCTTCAATTTCTCCTAGTTCAATCCGGAACCCAC^{GAATTTTCACCTGATCAAT}

Abl1-F

C.

Tb11-R

ATCGAATGGCCGCCGAGATCGA^{AAAA}AGTTATCTTTCACGCCTACTTGTTCTAACTTCAATAC
TTCTTTCCAGATGAGAACTAACTGCTCTTCAGTCTCGTTCTTGAGCAACATGGTCAATCC
CGGTTTGCATGCCTGCTGCGGGCAACGGTAATGCACTGCGATTACCTTTCCATTTGATGTT
AACGGCATTTGCTCCAGCTGCATGAGGTATGAAGGTATCATGTAGCCAGGTATTTGCTGAGC
AAGCTTTTCTTTGAATTGATTGATTGGCAGTGCTTGCTCCGCTACAAAATAAGCATACAAAT
CGCTTTGCCCTTCTGTGTTTTCTCTGGCCAGAACAACTGCTTCCTGTACATGCTCAATGTTG
AGAAGAGCGGCCTCGACCTCGTTTAGTTGATACGGTAACCGC^{GGATTTTCACTTGATGATC}

Abl1-F

D.

Abl1-F

GATCAACAGGTGAAAATT^{CG}GGGATTCCGGATTGAATTGGGAGAAATTGAAGCCGCGCTTGT
TCAGCATCCGAGCTTGAAGATGCGGCAGTGATCGTTCGTGAGGATCAGCCGGGAGATAAAC
GGTTAGCAGCTTATGTCATCCCTTCGGAAGAAACATTTGATACGGCGGAACTGCGCAGATAT
GCTGCTGAGAGACTGCCGGATTACATGGTGCCTGCTGCCTTTGTGACGATGAAAGAATTGCC
GCTGACACCGAATGGAAAGCTTGATCGTAAAGCGCTGCCCGCTCCGGATTTTCGAGCGGCAG
TGACAGGACGAGGACCGAGAACGCCGAGGAAGAAATTCTCTGTGACCTCTTTATGGAGGTG
CTTCACTTGCCGCGTGTGCGCATTGATGACCGTTTC^{TTTGATCTCGGCGGCCATTTCGAT}

Tb11-R

Figure 44. Séquences des fragments amplifiés par les amorces Abl1-F/Tb11-R (A. ADN S 45, B. ADN 1915, C. ADN 6633, D. ADN 168)

La taille des fragments amplifiés chez les souches de *B. thuringiensis* ne correspond pas à la taille des fragments déterminés *in silico* ce qui signifie que ces amorces n'ont pas détecté des gènes de bacillomycine.

L'identification de ces fragments a été effectuée par Blast N et Blast X dans la banque Genbank (NCBI). Le tableau 34 présente la nature des fragments amplifiés par PCR à l'aide des amorces Abl1-F et Tbl1-R et les prédictions des protéines correspondantes.

Tableau 34. Résultats Blast des amplicons obtenus par PCR avec les amorces Abl1-F et Tbl1-R

Souches	Taille des amplicons	BlastN Identité (%)	E. value	BlastX probable protéines
BT S 45	425 bp	NRPS à 94% ou 96% respectivement de <i>B. cereus</i> G9842 ou <i>B. cereus</i> ATCC 14579	0.00	Peptide synthétase de <i>B. thuringiensis</i> serovar <i>pulsiensis</i> BGSC 4CC1 ou <i>B. cereus</i> Rock1-15
BT 1915	425 bp	NRPS à 97% de <i>B. cereus</i> G9842 ou 94% à la gramicidine linéaire synthétase D de <i>B. cereus</i> B4264.	0.00	Peptide synthétase de <i>B. thuringiensis</i> serovar <i>berliner</i> ATCC 10792
BS 6633	434 bp	Mycosubtilin opéron à 99% de <i>B. subtilis</i> 6633 ou à 85% à BmyB de <i>B. amyloliquefaciens</i> FZB42	0.00	Mycosubtiline de <i>B. subtilis</i> ATCC 6633
BS 168	431 bp	<i>dhbF</i> gène à 99% de <i>B. subtilis</i> 168	0.00	Sidérophore 2,3-dihydroxybenzoate (DHB) de <i>B. subtilis</i> 168

Aucun gène de type bacillomycine n'a été détecté, ce qui indique l'absence de cette molécule dans les souches de *B. thuringiensis* S 45 et 1915. A priori, les souches de *B. thuringiensis* S 45 et 1915 ne contiennent pas de gènes de molécules de la famille des iturines car les deux couples d'amorces Am1-F/Tm1-R et Abl1-F/Tbl1-R n'ont détecté aucun fragment correspondant à des iturine synthétases. Ces amorces ont par ailleurs détecté un fragment similaire respectivement à 94% et 97% à une peptide synthétase de *B. thuringiensis* serovar *pulsiensis* ou *berliner* dans les deux souches étudiées.

Les deux fragments amplifiés par le couple Abl1-F/Tbl1-R respectivement chez *B. thuringiensis* 1915 et S 45 ont été comparés entre eux à l'aide de Needle Emboss (EBI, Cambridge, UK), ils ont montré un pourcentage d'identité égale à 96%.

La mise en évidence par les amorces Abl1-F/Tbl1-R de la mycosubtiline et d'un sidérophore à 99% avec une E. value = 0.00 respectivement chez les souches de *B. subtilis* ATCC 6633 et *B. subtilis* 168 confirme les travaux réalisés sur ces souches par certains auteurs (Duitman *et al.* 1999; Kunst *et al.* 1997).

Conclusion sur le criblage biochimique et moléculaire

Les analyses effectuées tant au niveau des activités antifongiques, qu'au niveau des analyses HPLC et Maldi-TOF et enfin au niveau moléculaire montrent que les deux souches de *Bacillus thuringiensis* S 45 et 1915 sont très proches. Ces résultats montrent la présence d'une molécule d'une masse identique à l'iturine C14 dans le surnageant de la souche S 45. Cependant, l'analyse par HPLC du surnageant de culture montre un pic avec un temps de rétention différent de l'iturine C14 et il n'a pas été possible de mettre en évidence les gènes codant pour des iturines synthétases. Par ailleurs, la mise en évidence d'un lipopeptide de masse similaire à celle des kurstakines chez *B. thuringiensis* S45 et la détection de gènes NRPS à la fois par les amorces Alb1-F/Tlb1-R chez les deux souches puis enfin par les amorces As1-F/Ts2-R chez la souche 1915 (Tableau 15) nous indiquent que ces souches synthétisent un lipopeptide NRPS.

Il s'avère très probable que les gènes NRPS détectés soient impliqués dans la synthèse de la molécule détectée et proche des kurstakines.

Il demeure toutefois d'être prudent car ces tests moléculaires ne permettent pas de corréliser le gène de cette peptide synthétase à la kurstakine détectée par Maldi-TOF car le gène qui coderait pour la kurstakine synthétase n'est pas connu (banques consultées en Août 2009). En effet, Hathout *et al.* qui ont découvert cette molécule ne l'ont pas décrite comme étant un lipopeptide synthétisé par voie non ribosomiale.

II. Analyses bioinformatiques des protéines détectées par les amorces spécifiques de lipopeptides

Les amorces As1-F/Ts2-R ont amplifié un fragment de 416 bp similaire (88%, E. value = 8e-130) à celui de l'acide adénylation chez *B. thuringiensis serovar berliner* 1915 (voir tableau 15). Cette synthétase serait responsable de la synthèse d'un peptide non ribosomal. Le pourcentage d'identité et la « E. value » montrent que la similarité entre ces deux fragments n'est pas parfaite, donc cela ne permet pas de confirmer la présence de ce gène.

En revanche, les amorces Abl1-F/Tbl1-R ont amplifié deux fragments respectivement chez *B. thuringiensis* 1915 et S 45 similaire à plus de 94% (E. value = 0.00) à une peptide synthétase. L'homologie élevée semble indiquer que ce gène est bien présent dans les souches étudiées.

Cette peptide synthétase est décrite comme étant une synthétase non ribosomiale de *Bacillus cereus* G9842 (numéro d'accension : ACK94377, Dodson *et al.* 2008; Read *et al.* 2009, non publié). L'analyse de cette peptide synthétase montre qu'elle est composée de 2156 aa formant deux modules. La recherche des acides aminés activés par ces différents modules par le logiciel NRPS Predictor a montré que cette protéine active la glutamine (70%) en position 1 et l'alanine

(100%) en position 2. Ces résultats sont confirmés par le logiciel NRPS/PKS (Ansari *et al.* 2004) qui précise que l'alanine située en position 2 serait de conformation D.

L'exploration du génome de *Bacillus cereus* G9842 où a été détecté la peptide synthétase, a montré la présence de deux autres phases de lecture ouverte codant des NRPS et annotées: synthétase de la gramicidine C (1518 aa, numéro d'accèsion : ACK95097) et synthétase de la bacitracine 1 (4960 aa, numéro d'accèsion : ACK95336). L'analyse de la structure de cette synthétase de la gramicidine C montre qu'elle est composée d'un module unique de type NRPS alors que la synthétase de la bacitracine 1 est constituée de 4 modules.

Selon le logiciel NRPS Predictor, le module unique de la gramicidine active une Thr (100%), les deux premiers modules de la synthétase de la bacitracine activent respectivement une Ser (100%) et une Pro (70%), les deux derniers incorporent deux molécules de Gln respectivement à 100% et 90%.

Le logiciel NRPS/PKS montre que la synthétase de la gramicidine C incorpore une Thr alors que celle de la bacitracine 1 incorporerait successivement les acides aminés suivants: Ser, Tyr, D-Gln et Gln.

Les différents outils de prédiction, à quelques nuances près, corroborent les résultats obtenus par l'un ou l'autre des logiciels.

Une des hypothèses probable est que ces 3 ORFs constituent un opéron impliqué dans la synthèse d'une molécule non ribosomiale. En effet, les 3 synthétases (gramicidine C, NRPS et bacitracine 1) seraient responsables de la synthèse d'une molécule unique.

Les amorces Abl1-F/Tbl1-R auraient mis en évidence un fragment de cette molécule présente dans les souches étudiées.

La séquence peptidique de cette molécule serait alors: Thr-Gln-D-Ala-Ser-Tyr(Pro)-D-Gln-Gln.

La figure 45 présente une proposition de la configuration de l'opéron responsable de cette molécule ainsi que les acides aminés activés obtenus à l'aide des logiciels NRPS/PKS et NRPS Predictor.

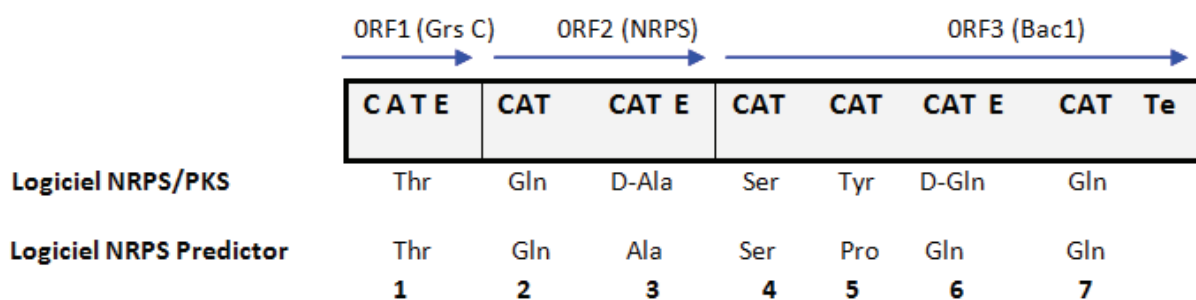


Figure 45. Proposition d'un schéma de l'opéron responsable de la molécule détectée chez *B. thuringiensis* serovar *berliner* à partir des gènes de la souche *B. cereus* G9842

L'introduction de cette composition dans la base de données NORINE ne permet pas de détecter un peptide NRPS connu. En comparant les acides aminés constitutifs de la séquence de la molécule avec tous les lipopeptides existants de *Bacillus*, on remarque qu'elle est très proche de celle des kurstakines. La figure 46 présente la structure de la chaîne peptidique des kurstakines.

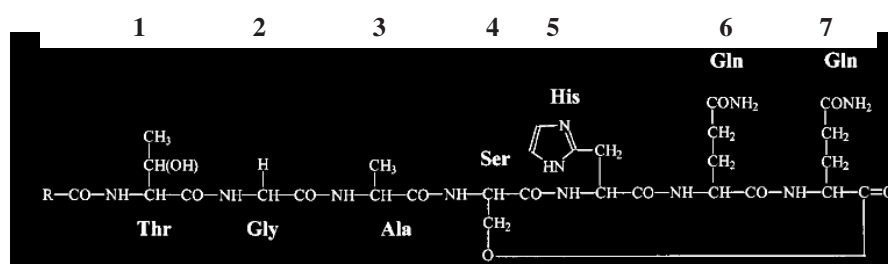


Figure 46. Structure de la kurstakine de *B. thuringiensis* *kurstaki* HD-1 (Hathout *et al.* 2000).

R= (CH₃)₂-H-- (CH₂)₇- masse moléculaire de 877.5 Da
R= CH₃-(CH₂)₁₀- masse moléculaire de 891.5 Da
R= (CH₃)₂-CH-(CH₂)₈- masse moléculaire de 891.5 Da
R= (CH₃)₂-H-- (CH₂)₉- masse moléculaire de 905.5 Da

La séquence en acides aminés de la molécule prédite diffère de celle des kurstakines par les acides aminés en position 2 et 5 (figure 46). En effet, la glycine des kurstakines est remplacée par une glutamine et l'histidine par une tyrosine ou une proline.

Cette différence pourrait être le fait des marges d'erreurs liées aux outils de prédictions.

Les deux outils (NRPS/PKS et NRPS Predictor) prédisent le même acide aminé en position 2 (Gln), celui en position 5 est reconnu comme une Pro avec une faible homologie par le logiciel NRPS Predictor alors que le logiciel NRPS/PKS l'identifie comme étant une Tyr.

Nous avons alors analysé la différence de prédiction des deux logiciels par l'analyse des signatures protéiques du site actif du domaine d'adénylation concerné.

Le tableau 35 présente les signatures protéiques du site actif du domaine A codant l'acide aminé en position 5 selon les logiciels NRPS Predictor et NRPS/PKS.

Tableau 35. Signatures protéiques des sites actifs de la protéine codant l'acide aminé en position 5 de la kurstakine (protéine cible)

	Selon NRPS Predictor	Selon NRPS/PKS
*Code NRPS pour l'acide aminé prédit	DALSVGEVVK	DSQFIGQVVK
Code NRPS du domaine d'adénylation étudié	DVQFIAHVVK	DSQSIGQVVK
Acide aminé prédit	Pro	Tyr

* Signature protéique du site actif présent dans la base de données du code Stachelhaus

Pour le même domaine d'adénylation, les deux logiciels trouvent deux signatures différentes pour le site actif et les comparent aux signatures du code Stachelhaus. Celle obtenue par le logiciel NRPS Predictor est similaire à 70% à la signature protéique codant la proline alors que celle obtenue par le logiciel NRPS/PKS a une homologie de 90% à celle de la Tyr. La différence au niveau de l'acide aminé prédit par les deux logiciels est le fait de leur fonctionnement différent pour la détermination de la signature protéique du site actif d'un domaine A.

On remarque cependant que le pourcentage de similarité est plus élevé pour le logiciel NRPS/PKS que celui obtenu par le logiciel NRPS Predictor. Cela pourrait s'expliquer par le fait que le logiciel NRPS/PKS (basé sur la comparaison entre la séquence cible et 2223 domaines différents et plusieurs structures de molécules) permet d'avoir une meilleure prédiction pour tous les acides aminés par rapport au logiciel NRPS Predictor qui lui est basé sur la comparaison de la séquence cible de 1230 domaines A (§ 6.2.1. et 6.2.2 du Chapitre 1).

Néanmoins, ces différentes prédictions montrent la présence de la kurstakine ou d'une molécule proche dans les souches de *B. thuringiensis* testées.

Conclusions et Perspectives

Ces analyses bioinformatiques confirment les résultats des analyses Maldi-TOF réalisées sur ces souches. En effet, l'analyse comparative des masses de la molécule détectée par Maldi-TOF en milieu solide de S 45 avec celles des kurstakines (878,47; 892,41; 906,51) mises en évidence par Hathout *et al.* (2000), la composition en acides aminés des chaînes peptidiques respectives, les

activités antifongiques observées chez les deux souches étudiées comparées à celles des kurstakines (anti-*Stachybotrys charatum*) montrent que la molécule détectée est une kurstakine.

Les kurstakines sont des lipopeptides qui ont été déjà décrits en outre comme possédant une activité antifongique. Leur chaîne acyl gras est reliée à l'extrémité N-terminale d'un résidu d'acide aminé via une liaison amide. Chaque molécule a une liaison lactone entre l'acide aminé C-terminal et le groupement hydroxyl du résidu sérine. Selon Hathout *et al.* (2000), les kurstakines sont composées de 4 molécules qui diffèrent uniquement sur la chaîne d'acides gras.

La structure de la kurstakine est proche de celle des autres lipopeptides synthétisés par les espèces de *Bacillus* (exemple des surfactines). Cependant, sa composition en acides aminés présente un résidu His, qui est absent dans les lipopeptides de *Bacillus* (Tableau 2).

Cette étude montre que les souches de *B. thuringiensis* serovar *berliner* 1915 et S 45 produisent un métabolite secondaire de type NRPS: une kurstakine.

Ces travaux soulignent pour la première fois que les kurstakines sont des molécules non ribosomiales et proposent un schéma de l'organisation de l'opéron qui serait impliqué dans leur synthèse.

Bien que cette étude permette de classer la molécule parmi dans la famille des kurstakines (lipopeptide de *Bacillus*), il est toutefois nécessaire que des travaux supplémentaires soient réalisés pour déterminer de façon expérimentale la composition en acides aminés, le séquençage complet de l'opéron et d'autres tests d'activités biologiques pour déterminer la kurstakine détectée.

Discussion générale

L'objectif principal de ce travail était la mise au point d'une technique de sélection à l'aide d'outils moléculaires de souches potentiellement productrices de molécules non ribosomiales. La littérature montre en effet de nombreux exemples de gènes codant pour des NRPS pas ou peu exprimés ou encore des synthétases non fonctionnelles. La sélection basée uniquement sur la détection de la présence d'un peptide NRPS dans le surnageant de culture d'une souche par des techniques de spectrométrie de masse, par exemple, est donc souvent limitée. Différentes stratégies ont déjà été proposées dans la littérature pour amplifier des gènes codant pour des NRPS, elles sont principalement basées sur la définition d'amorces non dégénérées ou dégénérées sur une base protéique. L'originalité de notre travail se situe dans un premier temps dans l'utilisation des gènes comme base de définition des amorces dégénérées et l'exploitation des séquences des domaines d'adénylation et de thiolation pour améliorer la sélectivité des amorces utilisées.

Le choix de la réaction en chaîne de la polymérase, technique très sensible, génère parfois des difficultés sur l'interprétation des résultats obtenus.

En effet, en l'absence de gènes cibles, la nature dégénérée des amorces peut conduire à l'amplification de gènes non NRPS. Parfois, l'obtention d'un amplicon résulte de l'hybridation d'une seule et même amorce. Différents critères d'analyse sont donc indispensables pour détecter les vrais résultats positifs. Tout d'abord l'observation, sur des gels d'électrophorèse, de l'intensité des bandes et de l'estimation de la taille des fragments par rapport à un marqueur est une indication qui peut permettre un premier tri sur le criblage d'une collection. Ensuite le séquençage du fragment amplifié et la recherche de son identité par Blast ou la comparaison avec d'autres gènes déjà existants et présents dans une banque d'ADN permettent une interprétation fiable de résultats PCR.

Les différences au niveau nucléique des synthétases impliquées dans la synthèse de molécules non ribosomiales, n'ont pas permis la mise au point d'un couple d'amorce capable d'amplifier par PCR tous les gènes NRPS.

Une nouvelle stratégie alors basée sur la détection de familles de molécules au sein d'espèces ciblées a été mise au point. Dans un premier temps, les analyses comparatives au niveau nucléique et protéique, des gènes et des synthétases qu'ils codent, respectivement, ont permis de montrer que des protéines très similaires telles que la fengycine et la plispastatine synthétase (85% d'homologie) sont codées par des gènes ne présentant que 42% d'homologie. De plus les gènes de la plipastatine sont plus proches de ceux de la surfactine, une molécule appartenant à

une autre famille, que de ceux codant pour la fengycine. Enfin les gènes des bacillomycines sont moins similaires (environ 42%) aux gènes codant pour les synthétases de molécules de la même famille (iturines et mycosubtilines) qui ont une homologie entre eux de plus de 70%. Cette stratégie a permis de détecter des lipopeptides au sein de souches du genre *Bacillus*.

Les différences des nucléotides observées au niveau des molécules pourraient s'expliquer par la grande flexibilité de lecture du 3^{ème} codon dans un triplet nucléotidique ou des préférences de lecture des codons au niveau de chaque espèce. Certains auteurs l'attribueraient à la dégénérescence du code génétique. En effet, la nature dégénérée du code génétique fait que la plupart des acides aminés peuvent être codés par plusieurs codons (triplets de nucléotides), que l'on appelle codons synonymes. Grantham *et al.* (1980) ont montré que les codons synonymes ne sont pas utilisés à la même fréquence, et que chaque espèce a sa propre signature de l'usage du code.

Cette première stratégie de criblage a conduit à un autre résultat intéressant. Elle a permis de révéler la présence du gène de la plipastatine chez *Bacillus subtilis* ATCC 21332. Jusqu'il y a peu, cette souche était considérée, dans la littérature, comme productrice uniquement de surfactine. Des travaux récents réalisés dans le laboratoire ProBioGEM par le Dr Gancel (Gancel *et al.* 2009) et par le Dr François Coutte (2009) ont mis en évidence de la plipastatine dans différents surnageants de culture de cette souche. Nos travaux apportent donc pour la première fois, la preuve moléculaire de la présence des gènes codant pour la plipastatine synthétase chez *B. subtilis* ATCC 21332.

Au vu des résultats générés, cette approche pour la définition des amorces spécifiques a alors été automatisée au moyen d'un logiciel mis à la disposition de notre laboratoire et permettra le criblage de nouvelles molécules. Elle est actuellement étendue au laboratoire à la sélection de souches de cyanobactéries productrices de NRPS en collaboration avec une entreprise.

Utilisées pour le criblage de deux collections de souches (collection de souches de *Bacillus* isolées de produits fermentés du Burkina-Faso par le Dr Aly Savadogo, post-doctorant et une collection de souches de *Bacillus thuringiensis* d'Algérie isolées et identifiées par Ahmed Abderrahmani, Doctorant), ces amorces ont mis en évidence :

- des souches productrices de surfactines, pour la première fois, dans ce type de produits fermentés en complément des techniques de spectrométrie de masse. Il s'agira en perspectives de vérifier la présence de la surfactine dans le produit lui-même et d'évaluer les avantages liés à cette présence.
- la synthèse de kurstakines dans une souche de *B. thuringiensis* isolée de céréales.

Cette étude nous a permis également de confirmer que cette molécule de kurstakine est synthétisée par voie non ribosomiale et de proposer par une analyse bioinformatique, l'organisation de l'opéron qui code pour les kurstakine synthétases. Nos résultats sont en conformité avec ceux publiés très récemment par Bumpus *et al.* (2009) sur la détection d'homologues de kurstakines chez *Bacillus sp.* NK2018. Les kurstakines, en effet, constituent une nouvelle classe de lipopeptides proche de la famille des iturines et des fengycines par leurs activités antifongiques décrites dans la littérature mais aussi des surfactines par leur poids moléculaire et le nombre des acides aminés constituant leur chaîne peptidique.

Nous postulons alors que les lipopeptides non ribosomiaux de *Bacillus* ne sont pas classés en trois familles comme l'ont suggéré plusieurs auteurs et que nous avons décrits dans le chapitre 3 mais plutôt en quatre familles (surfactines, fengycines, iturines et kurstakines).

Il s'agira dans la suite des travaux sur cette molécule d'améliorer sa productivité soit par la recherche d'un milieu spécifique approprié soit par génie moléculaire pour déterminer les potentialités biologiques liées à cette molécule. Il est, par ailleurs, probable que la présence éventuelle de l'histidine (acide aminé chargé positivement) dans sa chaîne peptidique augmente son adsorption sur des membranes cytoplasmiques de cellules.

L'identification des molécules pendant ces travaux a été possible grâce à la combinaison de la stratégie moléculaire que nous avons mise au point avec des techniques d'identification classiques de spectrométrie et de chromatographie. Cela n'a pas toujours été possible, notamment lorsque le gène détecté est muet comme chez *Lactobacillus plantarum* WCSF1. En effet, un des inconvénients de cette approche comme toutes les techniques moléculaires similaires réside dans le fait que la taille des fragments séquencés ne permet pas de les identifier. Une solution pour déterminer l'identité du gène dont le fragment a été détecté serait de combiner des variants de la technique PCR: l'Inverse et la Long Range Polymerase Chain Reaction (LR-IPCR). Cette combinaison de techniques moléculaires de PCR connaît un essor important depuis qu'elle a permis à certains auteurs de caractériser des gènes codant pour des protéines impliquées dans le métabolisme de certaines maladies humaines. En effet, Domenighini *et al.* (1990) ont caractérisé le gène de l'hémagglutinine filamenteuse de *Bordetella pertussis*, une glycoprotéine responsable de la toux chez l'adulte lorsqu'elle est activée. Khairatkar-Joshi et Ahmao (1998), ont aussi, par cette technique, séquencé le gène GAD (Acide Glutamique Décarboxylase), responsable de la synthèse d'un antigène lors du développement primaire du diabète de type I chez l'homme.

Conclusions générales

Cette étude, par les résultats générés, contribue de façon significative à l'amélioration des connaissances sur les métabolites secondaires en particulier les molécules non ribosomiales par l'apport d'outils moléculaires pour leur détection et aussi par l'apport d'outils bioinformatiques pour accélérer les techniques de criblage.

Ces travaux ont alors permis de corriger les difficultés du criblage de molécules non ribosomiales dans les surnageants de culture lorsque celles-ci ne sont pas produites pour diverses raisons. La combinaison de cette stratégie moléculaire et des méthodes classiques pour l'identification de molécules s'avère être un atout majeur.

En perspectives, il s'agira d'approfondir la stratégie actuelle de définition des amorces en se basant sur la phylogénie des domaines d'adénylation et ce afin de définir des amorces capables d'amplifier des gènes codant pour des familles de molécules produites par des espèces proches.

En effet, ces travaux ont montré que les phénomènes de transferts horizontaux de gènes peuvent expliquer la différence des monomères activés par des synthétases pourtant proches au niveau nucléotidique et protéique. Cela est particulièrement le fait des domaines intergéniques.

D'autres travaux pourraient porter sur la détection spécifique de certains domaines :

- épimérisation pour détecter des peptides contenant des résidus de type D

En effet, la présence d'acides aminés de conformation D dans une séquence peptidique est une indication sur l'origine non ribosomiale de la molécule. La mise au point d'une stratégie moléculaire détectant les domaines d'épimérisation permettra de signaler la présence de molécules de type non ribosomiale lors d'un criblage

- condensation (starter) pour détecter des lipopeptides

Rausch *et al.* (2005) ont montré la spécificité du domaine C starter responsable de l'incorporation de l'acide gras dans la formation des lipopeptides NRPS par rapport aux autres domaines de condensation. Une étude actuelle sur la phylogénie de ces domaines C starter permettra à terme de mettre au point une stratégie moléculaire pour détecter de façon spécifique le potentiel de synthèse de lipopeptides.

En outre, une identification de la kurstakine détectée chez *B. thuringiensis* sera effectuée par le séquençage de l'opéron de sa synthétase. De plus, la détermination des activités biologiques par des tests *in vitro* et *in vivo* seront réalisées afin d'envisager une production à grande échelle.

Publications et Communications

- **A. Tapi, M. Chollet-Imbert, B. Scherens, Ph. Jacques** (2009) New approach for the detection of non ribosomal peptide synthetase genes in *Bacillus* strains by Polymerase Chain Reaction. App Microbiol Biotechnol. DOI 10.1007/s00253-009-2176-4.
- **Ph Jacques, F. Gancel, M. Chollet-Imbert, JS Guez, M. Béchet, S. Caboche, F. Coucheney, F. Coutte, A. Tapi, V. Léclère** (2007) La synthèse peptidique non-ribosomiale, source de biodiversité de composés actifs. Bull Soc Microbiol. 22 : 234-240.
- **Arthur Tapi.** Recherche de nouveaux lipopeptides chez *Bacillus*: un enjeu pour les industries. Doctorale Franco-Belges Sept 2009 (Poster + communication orale)
- **Jacques P, Froidevaux R, Chollet-Imbert M, Tapi A, Caboche S, Pupin M, Kucherov G, Dhulster P, Le Flem G, Vercaigne-Marko D, Leclère V.** Novel approaches to isolate antimicrobial peptides using bioinformatics and molecular tools. Second international symposium on antimicrobial peptides-food, veterinary, medical and novel applications. St Malo (France), communications, 17-19 June 2009.
- **A. Tapi, M. Chollet-Imbert, B. Scherens, Ph. Jacques,** 5th International conference on Gram-Positive Microorganisms/15th International Conference on Bacilli, San-Diego, California. June 14-18, 2009
- **Savado, M. Chollet-Imbert, A. Tapi, Ph. Jacques.** NRPS molecules producing strains isolated from Bikalga and Soumbala fermented condiments of Burkina Faso. St Malo (France), communications, 17-19 June 2009.

REFERENCES
BIBLIOGRAPHIQUES

Références bibliographiques

- Adams MR, Marteau P** (1995) On the safety of lactic acid bacteria from food. *Inter J Food Microbiol* **27**: 263-264.
- Aguirre M, Collins MD** (1993) Lactic acid bacteria and human clinical infection. *J Appl Bacteriol* **75**: 95–107.
- Ansari MZ, Yadav G, Gokhale RS, Mohanty D** (2004) NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucl Ac Res* **32**: W405-13.
- Ayuso-Sacido A, Genilloud O** (2005) New PCR primers for screening of NRPS and PKS-I systems in actinomycetes: Detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microbial Ecol* **49**: 10-24.
- Barrios-Llerena ME, Burja AM, Wright PC** (2007) Genetic analysis of polyketide synthase and peptide synthetase genes in cyanobacteria as a mining tool for secondary metabolites. *J Ind Microbiol Biotechnol* **34**: 443-56.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL** (2008) Genbank. *Nucl Ac Res* **36**: D25-30.
- Berdnikova TF, Shashkov AS, Katrukha GS, Lapchinskaya OA, Yurkevich NV, Grachev AA, Nifant'ev NE** (2009) The structure of antibiotic eremomycin B. *Bioorganicheskaya Khimiya* **35**: 550–556.
- Borchert S, Patil SS, Marahiel MA** (1992) Identification of putative multifunctional peptide synthetase genes using highly conserved oligonucleotide sequences derived from synthetases. *FEMS Microbiol Lett* **92**: 175-180.
- Brahim S, Bettaieb A, Kenani A** (2008) Deglycosylated bleomycin triggers apoptosis in laryngeal carcinoma cells in a caspase and reactive oxygen species independent manner. *J Oral Pathol Med* **37**: 352-357.

Broadhurst RW, Nietlispach D, Wheatcroft MP, Leadlay PF, Weissman KJ (2003) The structure of docking domains in modular polyketide synthases. *Chem Biol* **10**: 723-731.

Bumpus SB, Evans BS, Thomas P M, Ntai I, Kelleher NL (2009) A proteomics approach to discovering natural products and their biosynthetic pathways. *Nature Biotech* **27**: 951-956.

Burgess C, O'Connell-Motherway M, Sybesma W, Hugenholtz J, Sinderen D V (2004) Riboflavin production in *Lactococcus lactis*: potential for in situ production of vitamin-enriched foods. *Appl Environ Microbiol* **70**: 5769–5777.

Byford MF, Baldwin JE, Shiau CY, Schofield CJ (1997) The mechanism of ACV synthetase. *Chem Rev* **7**: 2631-2650.

Caboche S, Pupin M, Leclère V, Jacques Ph, Kuchеров G (2008) Structural pattern matching of non-ribosomal peptides. *BMC Struct Biol* **9**: 15- 24.

Challis GL, Ravel J, Townsend CA (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**: 211-224.

Chang YH, Shangkuan YH, Lin HC, Liu HW (2003) PCR assay of the *groEL* gene for detection and differentiation of *Bacillus cereus* group cells. *Appl Environ Microbiol* **69**: 4502–4510.

Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsis A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Süßmuth RD, Borriss R (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol* **188**: 4024–4036.

Chollet-Imbert M, Gancel F, Slomianny C, Jacques P (2009) Differentiated pellicle organization and lipopeptide production in standing culture of *Bacillus subtilis* strains. *Arch Microbiol* **191**: 63–71.

Clugston SL, Sieber SA, Marahiel MA, Walsh CT (2003) Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a DCL catalyst. *Biochem* **42**: 12095-12104.

Coque JJ, Martin JF, Calzada JG, Liras P (1991) The cephamycin biosynthetic genes *pcbAB*, encoding a large multidomain peptide synthetase, and *pcbC* of *Nocardia lactamdurans* are clustered together together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. *Mol Microbiol* **5**: 1125–1133.

Conti E, Stachelhaus T, Marahiel MA, Brick P (1997) Structural basis for the activation of phenylalanine in the non ribosomal biosynthesis of gramicidin. *Embo J* **16**: 4174–4183.

Cosmina P, Rodriguez F, De Ferra F, Grandi G, Perego M, Venema G, Sinderen VD (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol Microbiol* **8**: 821–831.

Coutte F Production sélective des lipopeptides par *Bacillus subtilis* en bioréacteur en membranes. Thèse de Doctorat-Lille 1, 243 pp, soutenue en Sept 2009.

Daly NL, Love S, Alewood PF, Craik DJ (1999) Chemical synthesis and folding pathways of large cyclic polypeptides: studies of the cystine knot polypeptide kalata B1. *Biochem* **38**: 10606–10614.

Daubin V, Moran NA, Ochman H (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* **301**: 829–832.

De Crécy-Lagard V, Blanc V, Gil P, Naudin L, Lorenzon S, Famechon A, Bamas- Jacques N, Crouzet J, Thibaut D (1997) Pristinamycin I biosynthesis in *Streptomyces pristinaespiralis*: molecular characterization of the first two structural peptide synthetase genes. *J Bacteriol* **3**: 705-713.

De Crécy-Lagard V, Marlière P, Saurin W (1995): Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *CR Acad Sci III* **318**: 927-936.

Deleu M, Paquot M, Jacques Ph, Thonart P, Adriaensen Y, Dufrêne YV (1999) Nanometer scale organization of mixed surfactin/phosphatidylcholine monolayers. *Biophys J* **77**: 2304-2310.

Deleu M, Bouffieux O, Razafindralambo H, Paquot M, Hbid C, Thonart P (2003) Interaction of surfactin with membranes: a computational approach. *Langmuir* **19**: 3377–3385.

Dertz EA, Jide X, Stintzi A, Raymond KN (2006) Bacillibactin-mediated iron transport in *Bacillus subtilis*. *J Am Chem Soc* **126**: 22-23.

Doekel S, Coëffet-Le Gal MF, Gu JQ, Chu M, Baltz R H, Brian P (2008) Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiol* **154**: 2872-2880.

Du L, Sanchez C, Shen B (2001) Hybrid peptide-polyketide natural products: Biosynthesis and prospects toward engineering novel molecules. *Metab Eng* **3**: 78-95.

Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC 6633: A multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *PNAS* **96**: 13294-13299.

Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT (2000) The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *PNAS* **97**: 2509–2514.

Eppelmann K, Stachelhaus T, Marahiel MA (2002) Exploitation of the selectivity-conferring code of non ribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochem* **41**: 9718–9726.

Feignier C, Besson F, Michel G (1996) Characterization of iturin synthetase in the wild-type *Bacillus subtilis* strain producing iturin and in an iturin deficient mutant. *FEMS Microbiol Lett* **136**: 117 – 122.

- Fickers P, Guez JS, Damblon C, Leclère V, Béchet M, Jacques Ph, Joris B** (2009) High-level biosynthesis of the anteiso-C₁₇ isoform of the antibiotic mycosubtilin in *Bacillus subtilis* and characterization of its candidacidal activity. *Appl Environ Microbiol* **75**: 4636-4640.
- Firn RD, Jones CG** (2003) Natural products-a simple model to explain chemical diversity. *Nat Prod Rep* **20**: 382-391.
- Gancel F, Montastruc L, Liu T, Zhao L, Nikov I** (2009) Lipopeptide overproduction by cell immobilization on iron-enriched light polymer particles. *Process Biochem* **44**: 975-978.
- Gehring AM, Mori I, Walsh C T** (1998) Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochem* **37**: 2648-2659.
- Giacomodonato MN, Pettinari J, Souto GI, Meandez SB, Loapez NI** (2001) A PCR-based method for the screening of bacterial strains with antifungal activity in suppressive soybean rhizosphere. *World J Microbiol Biotechnol* **17**: 51-55.
- Ginolhac A, Jarrin C, Robe P, Perrie G, Vogel TM, Simonet P, Nalin R** (2005) Type I polyketide synthases may have evolved through horizontal gene transfer. *J Mol Evol* **60**: 716-725.
- Gokhale RS, Sankaranarayanan R, Mohanty D** (2007) Versatility of polyketide synthases in generating metabolic diversity. *Curr Opin Struct Biol* **17**: 736-743.
- Grangemard I, Wallach J, Maget-Dana R, Peypoux F** (2001) Lichenysin. A more efficient cation chelator than surfactin. *Appl Biochem Biotechnol* **90**: 199-210.
- Grantham R, Gautier C, Gouy M** (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucl Ac Res* **8**: 1892-1912.
- Grau A, Fernández JCG, Peypoux F, Ortiz A** (1999) A study on the interactions of surfactin with phospholipid vesicles. *Biochim Biophys Acta* **1418**: 307-319.

Hachmann AB, Angert E R, Helmann D J (2009) Genetic analysis of factors affecting susceptibility of *Bacillus subtilis* to daptomycin. *Antimicrob Agents Chemother* **53**: 1598-1609.

Hathout Y, Ho YP, Ryzhov V, Demirev P, Fenselau C (2000) Kurstakins: a new class of lipopeptides isolated from *Bacillus thuringiensis*. *J Nat Prod* **63**: 1492–1496.

Haavik HI (1974) Studies on the formation of bacitracin by *Bacillus licheniformis*: Effect of inorganic phosphate. *J Gen Microbiol* **84**: 226-230.

Heerklotz H, Seelig J (2001) Detergent-like action of the antibiotic peptide surfactin on lipid membranes. *Biophys J* **81**: 1547–1554.

Hoppert M, Gentzsch C, Schörgendorfer K (2001) Structure and localization of cyclosporin synthetase, the key enzyme of cyclosporin biosynthesis in *Tolypocladium inflatum*. *Arch Microbiol* **4**: 285-293.

Hopwood DA (2004) Cracking the polyketide code. *PLoS Biol* **2**: E35.

Hori K, Yamamoto Y, Minetoki T, Kurotsu T, Kanda M, Miura S, Okamura K, Furuyama J, Saito Y (1989) Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. *J Biochem* **106**: 639–645.

Hu LB, Shi ZQ, Zhang T, Yang ZM (2007) Fengycin antibiotics isolated from B-FS01 culture inhibits the growth of *Fusarium moniliforme* Sheldon ATCC 38932. *FEMS Microbiol Lett* **272**: 91-98.

Hubbard BK, Walsh CT (2003) Vancomycin assembly: nature's way. *Angew Chem Int* **42**: 730–765.

Jacques Ph, Hbid C, Destain J, Razafindralambo H, Paquot M, Pauw E, Thonart P (1999) Optimization of biosurfactant lipopeptide production from *Bacillus subtilis* S499 by Plackett-Burman design. *Appl Biochem Biotechnol* **77**: 223-233.

Jenke-Kodama H, Sandmann A, Muller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* **22**: 2027-2039.

Jullien N (2007) AmplifX software 1.4.4 Available via <http://ifrjr.nord.univ-mrs.fr/AmplifX> Home-page. Consulté depuis Nov 2006.

Kac G, Roux P, Poirot JL, Meyohas Mc, Cadranel J, Chouaid C, Watine J, Fourest F, Lancastre F (1995) *Aspergillus* and aspergillosis: a retrospective study in two Paris hospitals. *J Myco Med* **5**: 75-85.

Khairatkar-Joshi N, Ahmao NS (1998) Inverse polymerase chain reaction mediated chromosome walking within the human glutamic acid decarboxylase gene. *J Biosci* **23**: 265-269.

Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA (2004) The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J Biol Chem* **279**: 7413–7419.

Kleerebezem M et al. (2003) Complete genome sequence of *Lactobacillus plantarum* WCFS1. *PNAS* **100**: 1990-1995.

Kluge B, Vater J, Salnikow J, Eckart K (1988) Studies on the biosynthesis of surfactin, a lipopeptide antibiotic from *Bacillus subtilis* ATCC 21332. *FEBS Lett* **231**: 107-110.

Koltay A, Daly NL, Gustafson KR, Craik DJ (2004) Structure of circulin B and implications for antimicrobial activity of the cyclotides. *Int J Pep Res Therapeutics* **11**: 99-106.

Konz D, Doekel S, Marahiel M A (1999) Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *Am Soc Microbiol* **181**: 133-140.

Konz D, Klens A, Schörgendorfer K, Marahiel M A (1997) The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multimodular peptide synthetases. *Chem Biol* **4**: 927-937.

Kracht M, Rokos H, Ozel M, Kowall M, Pauli G, Vater J (1999) Antiviral and hemolytic activities of surfactin isoforms and their methyl ester derivatives. *J Antibiot* **52**: 613-619.

Kunst F et al. (1997) The complete genome sequence of the Gram-positive model organism *Bacillus subtilis* (strain 168). *Nature* **390**: 249-25.

Lawen A, Zocher R (1990) Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described. *J Biol Chem* **265**: 11355-11360.

Leclère V, Béchet M, Adam A, Guez JS, Wathelet B, Ongena M, Thonart P, Gancel F, Chollet-Imbert M, Jacques Ph (2005) Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. *Appl Environ Microbiol* **71**: 4577-84.

Lee YK, Kim SB, Park CS, Kim JG, Oh HM, Yoon BD, Kim HS (2005) Chromosomal integration of *sfp* gene in *Bacillus subtilis* to enhance bioavailability of hydrophobic liquids. *Appl Microbiol Biotechnol* **67**: 789-794.

Lin TP, Chen CL, Chang LK, Tschen JSM, Liu ST (1999) Functional and transcriptional analyses of a fengycin synthetase gene, *fenC*, from *Bacillus subtilis*. *J Bacteriol* **181**: 5060-5067.

Magnusson SP, Qvortrup K, Larsen JO, Rosager S, Hanson P, Aagaard P, Krogsgaard M & Kjaer M (2002) Collagen fibril size and crimp morphology in ruptured and intact Achilles tendons. *Matrix Biol* **21**: 369–377.

Marahiel M A, Stachelhaus T, Mootz H D (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* **97**: 2651–2674.

Markowitz et al (2007) The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucl Ac Res* 1-6.

Miller DA, Walsh CT, Luo L (2001) C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase. *J Am Chem Soc* **123**: 8434-5.

- Molenaar D, Bringel F, Schuren FH, de Vos WM, Siezen RJ, Kleerebezem M** (2005) Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J Bacteriol* **187**: 6119–6127.
- Mofredj A, Bahloul H, Chanut C** (2007) *Lactococcus lactis*: an opportunistic bacterium? *Med Mal Infect* **37**: 200–7.
- Mootz HD, Marahiel MA** (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* **179**: 6843–6850.
- Mootz HD, Schwarzer D, Marahiel MA** (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chem Biochem* **3**: 490–504.
- Morikawa M, Hirata Y, Imanak T** (2000) A study on the structure–function relationship of the lipopeptide biosurfactants. *Biochim Biophys Acta* **1488**: 211–218.
- Moss SJ, Martin CJ, Wilkinson B** (2004) Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat Prod Rep* **21**: 575–593.
- Moyne AL, Cleveland TE, Tuzun S** (2004) Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. *FEMS Microbiol Lett* **234**: 43–49.
- Moyne AL, Shelby R, Cleveland TE, Tuzun S** (2001) Bacillomycin D: an iturin with antifungal activity against *Aspergillus flavus*. *J Appl Microbiol* **90**: 622–629.
- Nagaraj G, Uma MV, Shivayogi MS, Balaram H** (2001) Antimalarial activities of peptide antibiotics isolated from fungi. *Antimicrob Agents Chemother* **45**: 145–149.
- Nakano MM, Marahiel MA, Zuber P** (1988) Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *B. subtilis*. *J Bacteriol* **170**: 5662–5668.
- Naruse N, Tenmyo O, Kobaru S, Kamei H, Miyaki T, Konishi M, Oki T.** (1990) Pumilacidin, a complex of new antiviral antibiotics. *J Antibiot* **43**: 267–280.

Needleman SB, Wunsch CD (1970) Needleman-Wunsch global alignment. *J Mol Biol* **48**: 443-453.

Neilan B, Dittmann E, Rouhiainen L, Bass RA, Schaub V, Sivonen K, Börner T (1999) Nonribosomal peptide synthesis and toxigenicity of Cyanobacteria. *J Bacteriol* **181**: 4089-4097.

Nettles JH, Li H, Cornett B, Krahn JM, Snyder JP, Downing KH (2004) The binding mode of epothilone A on tubulin by electron crystall. *Science* **305**: 866-305.

Nikolskaya AN, Panaccione DG, Walton JD (1995) Identification of peptide synthetase encoding genes from filamentous fungi producing host-selective phytotoxins or analogs. *Gene* **165**: 207-211.

Nishikiori T, Naganawa H, Muraoka Y, Aoyagi T, Umezawa H (1986) Plipastatins: new inhibitors of phospholipase A2, produced by *Bacillus cereus* BMG302-fF67. III. Structural elucidation of plipastatins. *J Antibiot* **39**: 755–761.

Ongena M, Duby F, Jourdan E, Beaudry T, Jadin V, Dommes J, Thonart P (2005a). *Bacillus subtilis* M4 decreases plant susceptibility towards fungal pathogens by increasing host resistance associated with differential gene expression. *Appl Microbiol Biotechnol* **67**: 692-698.

Ongena M, Jacques Ph (2008) *Bacillus* lipopeptides: Versatile weapons for plant disease biocontrol. *Trends Microbiol* **16**: 116-125.

Ongena M, Jacques Ph, Toure Y, Destain J, Jabrane A, Thonart P (2005b) Involvement of fengycin-type lipopeptides in the multifaceted biocontrol potential of *Bacillus subtilis*. *Appl Microbiol Biotechnol* **69**: 29-38.

Ovchinnikova TV, Murashev AN (2006) The peptaibol antibiotic zervamicin displays neurotropic activity. *Doklady Biochem Biophys* **414**: 707–709.

Parkouda C, Diawara B, Ouoba LI (2008) Technology and physico-chemical characteristics of Bikalga, alkaline fermented seeds of *Hibiscus sabdariffa*. *Afric J Biotechnol* **7**: 916–922.

Pelzer S, Reichert W, Huppert M, Heckmann D, Wohlleben W (1997) Cloning and analysis of a peptide synthetase gene of the balhimycin producer *Amycolatopsis mediterranei* DSM 5908 and development of a gene disruption/replacement system. *J Biotechnol* **56**: 115–128.

Peypoux F, Bonmatin JM, Wallach JM (1999) Recent trends in the biochemistry of surfactin. *J Appl Microbiol Biotechnol* **51**: 553-563.

Peypoux F, Guinand M, Michel G, Delcambe L, Das C, Lederer E (1978) Structure of iturin A, a peptidelipid antibiotic from *Bacillus subtilis*. *Biochem* **17**: 3992–3996.

Peypoux F, Lapr v te O, Pagadoy M, Wallach J (2004) *N*-Acyl derivatives of Asn, new bacterial *N*-acyl *D*-amino acids with surfactant activity. *Amino Acids* **26**: 209-214.

Pfeifer E, Pavela-Vrancic M, Doehren VH, Kleinkauf H (1995) Characterization of tyrocidine synthetase 1 (TY1): Requirement of posttranslational modification for peptide biosynthesis. *Biochem* **34**: 7450–7459.

Podlesek Z, Grabnar M (1987) Genetic mapping of the bacitracin synthetase gene(s) in *Bacillus licheniformis*. *J Gen Microbiol* **133**: 3093-3097.

Price NPJ, Rooney AP, Swezey JL, Perry E, Cohan FM (2007) Mass spectrometric analysis of lipopeptides from *Bacillus* strains isolated from diverse geographical locations. *FEMS Microbiol Letters* **271**: 83 – 89.

Psurek A, Neus  ss C, Pelzing M, Scriba GKE (2005) Analysis of the lipophilic peptaibol alamethicin by nonaqueous capillary electrophoresis-electrospray ionization-mass spectrometry. *Electrophoresis* **26**: 4368-4378.

Rajendran N (1999) Identification and cloning of a gene locus encoding peptide synthetase of *Pseudomonas fluorescens* by two sets of PCR primers. *Z Naturforschung* **54**: 105-109.

Ramarathnam R, Bo S, Chen Y, Dilantha WG, Xueven G, Kievit T (2007) Molecular and biochemical detection of fengycin and bacillomycin D-producing *Bacillus spp.*, antagonistic to fungal pathogens of canola and wheat. *Can J Microbiol* **53**: 901-911.

Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson D H (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using Transductive Support Vector Machines (TSVM). *Nucl Ac Res* **33**: 5799-5808.

Razafindralambo H, Popineau Y, Deleu M, Hbid C, Jacques Ph, Thonart P, Paquot M (1998) Foaming properties of lipopeptides produced by *Bacillus subtilis*: effect of lipid and peptide structural attributes. *J Agric Food Chem* **46**: 911-916.

Recktenwald J, Shawky R, Puk O, Pfennig F, Keller U, Wohlleben W, Pelzer S (2002) Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. *Microbiol* **148**: 1105–1118.

Rey et al (2004) Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol* **5**: r77.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365-86.

Rupf S, Merte K, Eschrich K (1999) Quantification of bacteria in oral samples by competitive polymerase chain reaction. *J Dent Res* **78**: 850–856.

Schauwecker F, Pfennig F, Schroder W, Keller U (1998) Molecular cloning of the actinomycin synthetase gene cluster from *Streptomyces chrysomallus* and functional heterologous expression of the gene encoding actinomycin synthetase II. *J Bacteriol* **180**: 2468–2474.

Schwarzer D, Marahiel MA (2001) Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* **88**: 93–101.

Schwarzer D, Finking R, Marahiel MA (2003) Nonribosomal peptides: from genes to products. *Nat Prod Rep* **20**: 275–287.

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**: 285-295.

Sieber SA, Marahiel MA (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105**: 715-738.

Siezen R J, Starrenburg MJC, Boekhorst J, Renckens B, Molenaar D, Vlieg JET (2008) Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl Environ Microbiol* **74**: 424–436.

Silakowski B, Kunze B, Nordsiek G, Blocker H, Hofle G, Müller R (2000) The myxochelin ion transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *Eur J Biochem* **267**: 6476.

Stachelhaus T, Marahiel MA (1995) Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J Biol Chem* **270**: 6163–6169.

Stachelhaus T, Mootz H D, Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**: 493-505.

Stachelhaus T, Walsh CT (2000) Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochem* **39**: 5775–5787.

Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**: 380-416.

Steller S, Sokoll A, Wilde C, Bernhard F, Franke P, Vater J (2004) Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochem* **43**: 11331-11343.

Steller S, Vollenbroich D, Leenders F, Stein T, Conrad B, Hofemeister J, Jacques Ph, Thonart P, Vater J (1999) Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chem Biol* **6**: 31–41.

Stryer L (1996) Heidelberg-Berlin-Oxford: Spektrum Akademischer Verlag GmbH. *Biochem* **4**.

Tang Y, Koppisch AT, Khosla C (2004) The acyltransferase homologue from the initiation module of the R1128 polyketide synthase is an acyl-ACP thioesterase that edits acetyl primer units. *Biochem* **43**: 9546-9555.

Tapi A, Chollet-Imbert M, Scherens B, Jacques Ph (2009) New approach for the detection of non ribosomal peptide synthetase genes in *Bacillus* strains by Polymerase Chain Reaction. *Appl Microbiol Biotechnol* (DOI 10.1007/s00253-009-2176-4).

Thomas DW, Ito T (1969) The revised structure of the peptide antibiotic esperin, established by mass spectrometry. *Tetrahedron* **25**: 1985–1990.

Tomino S, Yamada M, Itoh H, Kurahashik (1967) Cell-free synthesis of gramicidin S. *Biochem* **6**: 2552–2560.

Tor Y, Libman J, Shanzer A, Felder CE, Lifson S (1992) Chiral siderophore analogs: enterobactin. *J Americ Chem Soc* **114**: 6661-6671.

Trauger JW, Walsh CT (2000) Heterologous expression in *Escherichia coli* of the first module of the nonribosomal peptide synthetase for chloroeremomycin, a vancomycin-type glycopeptides antibiotic. *PNAS* **97**: 3112–3117.

Tsuge K, Akiyama T, Shoda M (2001) Cloning, sequencing, and characterization of the iturin A operon. *J Bacteriol* **183**: 6265-6273.

Tsuge K, Matsui K, Itaya M (2007) Production of the non-ribosomal peptide plipastatin in *Bacillus subtilis* regulated by three relevant gene blocks assembled in a single movable DNA segment. *J Biotechnol* **129**: 592-603.

Turgay K, Marahiel MA (1994) A general approach for identifying and cloning of peptide synthetase genes. *Peptide Res* **7**: 238-241.

Ullrich C, Kluge B, Palacz Z, Vater J (1991) Cell-free biosynthesis of surfactin, a cyclic lipopeptide produced by *Bacillus subtilis*. *Biochem* **30**: 6503–6508.

Umezawa H, Aoyagi T, Nishikiori T, Okuyama A, Yamagishi Y, Hamada M, Takeuchi T (1986) Plipastatins: new inhibitors of phospholipase A2, produced by *Bacillus cereus* BMG302-fF67. I. Taxonomy, production, isolation and preliminary characterization. *J Antibiot* **39**: 737-744.

Vater J, Kablitz B, Wilde C, Franke P, Mehta N, Cameotra SS (2002) Matrix-assisted laser desorption ionization-time of flight mass spectrometry of lipopeptide biosurfactants in whole cells and culture filtrates of *Bacillus subtilis* C-1 isolated from petroleum sludge. *Appl Environ Microbiol* **68**: 6210-6219.

Volpon L, Besson F, Lancelin JM (2000) NMR structure of antibiotics plipastatins A and B from *Bacillus subtilis* inhibitors of phospholipase A2. *FEBS Lett* **485**: 76–80.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189-1191.

Wattiau P, Renard M E, Ledent P, Debois V, Blackman G, Agathos SN (2001) A PCR test to identify *Bacillus subtilis* and closely related species and its application to the monitoring of wastewater biotreatment. *Appl Microbiol Biotechnol* **56**: 816-819.

Whitmore L, Wallace BA (2004) The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols. *Nucl Ac Res* **32**: D593-D594.

Wu XY, Walker MJ, Hornitzky M, Chin J (2005) Development of a group-specific PCR combined with ARDRA for the identification of *Bacillus* species of environmental significance. *J Microbiol Met* **64**: 107-119.

Wyckoff EE, Valle AM, Smith SL, Payne SM (1999) A multifunctional ATP-binding cassette transporter system from *vibrio cholerae* transports vibriobactin and enterobactin. *J Bacteriol* **181**: 7588-7596.

Zhang T, Shi ZQ, Hu LB, Cheng LG, Wang F (2008) Antifungal compounds from *Bacillus subtilis* B-FS06 inhibiting the growth of *Aspergillus flavus*. *World J Microbiol Biotechnol* 24:783–788

Zhang ZY, Liu C, Zhu YZ, Zhong Y, Zhu YQ, Zheng HJ, Zhao GP, Wang SY, Guo XK (2009) Complete genome sequence of *Lactobacillus plantarum* JDM1. *J Bacteriol* **191**: 5020–5021.

Zhou Z, Lai JR, Walsh CT (2007) Directed evolution of aryl carrier proteins in the enterobactin synthetase. *PNAS* **104**: 11621–11626.

Zuber P, Nakano MM, Marahiel MA (1993) Peptide antibiotics. In AL Sonenshein JA Hoch, and R Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: physiol, biochem, and mol biol. 897–916. American Society for Microbiology, Washington D.C.

ANNEXES

Annexes

Annexe 1: Publication 1

Article publié concernant le chapitre 6 sur la mise au point d'amorces dégénérées spécifiques au genre *Bacillus*.

Annexe 2 : Publication 2

Extrait Article soumis concernant l'utilisation d'outils moléculaires et de spectrométrie pour la détection des surfactines dans les souches isolées de Soumbala et de Bikalga

Annexe 3 : Séquence protéique de la kurstakine synthétase

Annexe 4 : Code génétique, Acides aminés et masse molaire (g/mol)

Annexe 5 : Quelques spectres Maldi-TOF

ANNEXE 1. PUBLICATION

Appl Microbiol Biotechnol
DOI 10.1007/s00253-009-2176-4

APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

New approach for the detection of non-ribosomal peptide synthetase genes in *Bacillus* strains by polymerase chain reaction

Arthur Tapi · Marlène Chollet-Imbert · Bart Scherens ·
Philippe Jacques

Received: 8 June 2009 / Revised: 29 July 2009 / Accepted: 30 July 2009
© Springer-Verlag 2009

Abstract *Bacillus* strains produce non-ribosomal lipopeptides that can be grouped into three families: surfactins or lichenysins, iturins and fengycins or plispastatins. These biosurfactants show a broad spectrum of biological activities. To detect strains able to produce these lipopeptides, a new polymerase chain reaction screening approach was developed using degenerated primers based on the intra-operon alignment of adenylation and thiolation nucleic acid domains of all enzymes implicated in the biosynthesis of each lipopeptide family. The comparative bioinformatics analyses of each operon led to the design of four primer pairs for the three families taking into account the differences between open reading frames of each synthetase gene. Tested on different *Bacillus* sp. strains, this technique was used successfully to detect not only the expected genes in the lipopeptide producing strains but also the presence of a plispastatin gene in *Bacillus subtilis* ATCC 21332 and a gene showing a high similarity with the polyketide synthase type I gene in the *B. subtilis* ATCC 6633 genome. It also led to the discovery of the presence of non-ribosomal peptide synthetase genes in *Bacillus thuringiensis* serovar

berliner 1915 and in *Bacillus cereus* LMG 2098. In addition, this work highlighted the differences between the fengycin and plipastatin operon at DNA level.

Keywords *Bacillus* sp. · Non-ribosomal peptide synthetase · PCR · Lipopeptide · Degenerated primers

Introduction

Non-ribosomal peptide synthetases (NRPS) are multi-modular enzymes that make atypical peptides through a thiotemplate mechanism independent of ribosomes. These peptides have a broad spectrum of biological activities and are used in medicine and biological research. They also have a lot of potential applications in the agricultural and environmental fields (Ongena and Jacques 2008). These compounds are produced by a large variety of microorganisms such as bacteria and fungi (Caboche et al. 2008).

Among them, the lipopeptides constitute a well-known class of amphiphilic membrane-active biosurfactant peptides with potent antimicrobial activities. Surfactin, fengycin and iturin families are representative of lipopeptide compounds produced by *Bacillus subtilis*. Surfactin and iturin compounds are cyclic lipopeptides, which contain a β -hydroxy fatty acid and a β -amino fatty acid respectively, as their lipophilic component (Bonmatin et al. 2003; Ongena and Jacques 2008). Fengycin is a lipodecapeptide with a β -hydroxy fatty acid in its side chain (Steller et al. 1999; Schneider et al. 1999). These agents can be classified in families of closely related isoforms, which differ in the length and branching of their fatty acid side chains and in the amino acid substitutions in the peptide rings (Ongena and Jacques 2008).

Electronic supplementary material The online version of this article (doi:10.1007/s00253-009-2176-4) contains supplementary material, which is available to authorized users.

A. Tapi · M. Chollet-Imbert · P. Jacques (✉)
Laboratoire des Procédés Biologiques, Génie Enzymatique et
Microbien, UPRES-EA 1026, Polytech'Lille/IUTA,
Université des Sciences et Technologies de Lille,
Av. Paul Langevin, Villeneuve d'Ascq,
59655 Lille, France
e-mail: philippe.jacques@polytech-lille.fr

B. Scherens
Institut de Recherches Microbiologiques JM Wiame,
Av. E. Gryson 1,
1070 Brussels, Belgium

Published online: 03 September 2009

 Springer

et al. 1986), iturin A and C (Tsuge et al. 2001), bacillomycin D, F and L (Moyne et al. 2004), bacillomycin Le (Eshita et al. 1995) and mycosubtilin (Leclère et al. 2005). Compared to chemical agents, lipopeptides are safer for the environment (Vanittanakom et al. 1986), and several reports have described their valuable role in the use of *Bacillus* strains as biocontrol agents of plant diseases (Ongena et al. 2005; Ongena and Jacques 2008; Chen et al. 2009).

However, the search for molecules in the supernatant of cultures does not always make it possible to highlight the capacity of synthesis of the bacteria either because the processes of culture implemented are not optimised or because the lack of an essential gene prevents the expression of this phenotype. Indeed, the presence of a fengycin lipopeptide was shown in *B. subtilis* S499, which had previously been characterised for its ability to produce only two families of lipopeptides: iturin A and surfactin S1 (Jacques et al. 1999; Ongena et al. 2005). Another well-known example is the case of *B. subtilis* 168, which shows in its genome, the complete and functional operons of surfactin and plipastatin but which is unable to produce either of these lipopeptides because of the presence of a frame-shifted *sfp* gene, which usually encodes a phosphopantetheinyl transferase responsible for the transformation of the apo form of synthetases in holo form (Lee et al. 2005; Tsuge et al. 2007). The introduction of this gene from *B. subtilis* ATCC 21332 in the JH 642 strain, a derivative of the standard *B. subtilis* 168, confirmed the production of surfactin lipopeptides (Nakano et al. 1988).

For these reasons, the study of the genetic potential could be useful to detect lipopeptide producer strains. The aim of this study is to set up a polymerase chain reaction (PCR) strategy for the screening of *Bacillus* strains that could be producers of non-ribosomal peptides and particularly lipopeptide compounds.

Several techniques or methods to design primers are commonly used to identify bacterial and fungal genes by PCR. The choice of the primers is the key element when screening for specific gene fragments. Until now, two strategies were used to detect NRPS genes. The first one defined degenerated PCR primers based on the consensus amino acid sequences inside NRPS domains from *B. subtilis* (Turgay and Marahiel 1994). In this method, two sets of primers based on highly conserved adenylation and thiolation regions of *B. subtilis* were used to identify and clone peptide synthetase genes in different *Bacillus* species (Giacomodonato et al. 2001; Wu et al. 2006) or other bacterial genera as *Pseudomonas* (Rajendran 1999). This method was also applied to other microorganisms such as filamentous fungi (Nikolskaya et al. 1995) and cyanobacteria (Neilan et al. 1999; Barrios-Ilerena et al. 2007). This strategy is based on back translation (reverse translation from protein to DNA) of the consensus sequence. When applied to

microorganisms other than the ones in which the consensus sequence was determined, codon bias was taken into account.

A second approach is based on the consensus region that was determined by aligning the nucleotide sequence of the adenylation domains of NRPS coding genes.

Out of a collection of 210 actinomycetes strains, Ayuso-Sacido and Genilloud (2005) specifically amplified NRPS sequences not only from the well-known productive *Streptomyces* but also from other genera not yet known to date as bioactive compound producers. More recently, Ramarathnam et al. (2007) screened bacteria (isolated from canola and wheat) that exhibit antifungal activity. They looked for the presence of fengycin D1 and bacillomycin D using two specific, non-degenerated PCR primers based on the adenylation domains of the corresponding genes. Because their primers were not degenerated and their design was limited to two modules from each operon, this method does not take into account the possible differences between synthetases in each *Bacillus*. It follows that the use of this method would not be appropriate for screening the genes of these peptides in all *Bacilli*. Moreover, other lipopeptides cannot be detected with this approach.

In this paper, we demonstrate the use of degenerated primers based on an alignment of domains inside *B. subtilis* synthetases involved in the synthesis of lipopeptide family. This technique will allow us to specifically screen for, amplify and clone surfactin, iturin or fengycin synthetase genes as well as other NRPS genes in different *Bacillus* strains.

Materials and methods

Bacterial strains, plasmids and culture conditions

The strains used in this work are presented in Table 1. *Bacillus* sp. and *Escherichia coli* strains were grown aerobically in Luria-Bertani (LB) medium. *Lactobacillus* strains were grown in de Man–Rogosa–Sharpe medium. The Petri-dish cultures used the same media containing 1.7% (w/v) agar (SO.B1.GEL, Hendaye, France). The strains were grown 24 h, at 30°C for *Bacillus* and *Lactobacillus* and at 37°C for *E. coli*.

For PCR experiments, two *Lactobacillus* strains explored by Molenaar et al. (2005) were used as a negative NRPS control because *Bacillus* genus was known to harbour NRPS.

DNA isolation

Genomic DNA was prepared using the Wizard Genomic DNA Purification Kit and protocol (Promega Corp.,

Table 1 Properties of bacteria used in this study

Strains	Genotype	Sequenced lipopeptide operon	Lipopeptide produced	Sources
<i>Bacillus subtilis</i> ATCC 21332	Wild type	Not determined	Surfactin	ATCC
<i>Bacillus subtilis</i> ATCC 9943	Wild type	Not determined	Surfactin, iturin A, fengycin	ATCC
<i>Bacillus subtilis</i> 168	<i>trpC2, sfp⁰</i>	Srf, Pli	No lipopeptide	<i>Bacillus</i> stock genetic Center
<i>Bacillus subtilis</i> ATCC 6633	Wild type	Myc	Surfactin, mycosubtilin	ATCC
<i>Bacillus licheniformis</i> ATCC 14580	Wild type	Lch	Lichenysin	ATCC
<i>Bacillus thuringiensis</i> berliner 1915	Wild type	Not determined	Not determined	CRBIP, Institut Pasteur, Paris, France
<i>Bacillus cereus</i> LMG 2098	Wild type	Not determined	Not determined	LMG, Ghent, Belgium
<i>Lactobacillus plantarum</i> ATCC 14917	Wild type	Not determined	No lipopeptide	LMG, Ghent, Belgium
<i>Lactobacillus plantarum</i> ATCC 8014	Wild type	Not determined	No lipopeptide	LMG, Ghent, Belgium
<i>Escherichia coli</i> JM 109	<i>recA1, endA1, gyrA96, thi, hsdR17</i> (rK ⁻ , mK ⁺), <i>relA1, supE44, Δ(lac-proAB)</i> , [F', <i>traD36, proAB, lacIqZΔM15</i>]			Promega Corp. Madison, USA

Madison, USA). DNA concentration was established using a Nanodrop™ 1000 Spectrophotometer. DNA was used directly as genomic DNA for PCR amplification or stored at -20°C for further utilisation.

Primer design

Known nucleic sequences of NRPS biosynthetic clusters characterised in *B. subtilis* strains that are involved in the synthesis of these three families of lipopeptides (surfactins, mycosubtilins and fengycins) were analysed with Needle online software (Needleman and Wunsch 1970) and ORF Finder (NCBI, USA). We designed four primer pairs based on the conserved motifs obtained after alignment of the nucleic acid sequences identified as adenylation (A) or

thiolation (T) domains using the Vector NTI Advance 10™ software (Invitrogen, Carlsbad, USA). It was decided to use the mathematical approach called “coefficient of hybridization” (HyC) to calculate the hybridization capacities of the primers. Using International Union of Pure and Applied Chemistry (IUPAC) code, a coefficient 3 or 2 was affected when the nucleotide code was B, D, H or V or R, Y, S, W, K or M, respectively. HyC was calculated by multiplying the different values of the nucleotides making up each primer. Only primers with a HyC value below 200 were retained because the hybridization of the oligonucleotides was affected beyond. Degenerated forward primers were As1-F, Am1-F, Ap1-F and Af2-F, and the reverse ones were Ts2R, Tm1-R, Tp1-R and Tf1-R. One example of primer design is illustrated in Fig. 2.

Fig. 2 Schematic representation of conserved motifs from the alignment of adenylation and thiolation domains from mycosubtilin synthetase gene used for the design of degenerated primers by Vector NTI 10™ Advance Software

	5'	3'	3'	5'
A2 Myc C	CATCAGGTGAAAAGTCCGAG G		T Myc A	CATTTCCTTCGATCTTGAGGGG
A2 Myc B	GAGCAAGTCAAAATCAGAGG		T3 Myc B	GACTTTTTTGATATCGGTGG
A Myc A	GAACAAGTCAAAATCCGAGG		T1 Myc B	GATTTCTTTGATTTTGCGGG
A4 Myc B	CATCAAGTCAAAATTCGCGG		T2 Myc B	GATAATTTCTTTGATCTTGG
A1 Myc B	CATCAAGTGAAAATCCGCGG		T1 Myc C	CATCAATTGTTTGAGGAACA
A1 Myc C	CACCAGGTGAAAATTCGGGG		T2 Myc C	GATAACTTCTTTGACGTTGG
A3 Myc B	CATCAGGTGAAAATTCATG G		T4 Myc B	GATAACTTTTTTCGATCTCGG
Am1-F	CAKCAAGTSAAAATYCG MG G		Tm1-R	CCDASATCAAARAADTTATC

Another primer pair Prbac1/Prbac2 (Rupf et al. 1999) reported to be eubacteria-specific PCR primers was used to ensure the quality of the template DNA in the PCR. The oligonucleotide primers were ordered commercially (Eurogentec S.A, Liege, Belgium). Primer sequence was tested against numerous DNA sequences available in GenBank databases with AmpliT_X 1.4.4 Software (Jullien 2007) to determine their theoretical properties and specificities.

PCR conditions

The PCR conditions were experimentally determined and were different for each pair of primers. To find the correct annealing temperatures for each pair of primers, a PCR using an annealing temperature gradient was carried out in an Eppendorf AG (Hamburg, Germany) Mastereycler gradient. The amount of crude DNA extract in each PCR mixture was between 10 and 30 ng μL^{-1} .

The PCR thermal cycle programme included an initial denaturation at 94°C for 3 min, followed by 30 cycles, with a denaturation step at 94°C for 1 min, an annealing step of 30 s, at 43°C with As1-F/Ts2-R, at 45°C with Am1-F/Tm1-R and Af2-F/Tf1-R and at 58°C with Ap1-F/Tp1-R, followed by extension step during 45 s at 72°C except with Ap1-F/Tp1-R primers (75 s at 72°C). Final extension was performed at 72°C for 10 min. Annealing temperature for the Prbac1/Prbac2 primers used as positive control was identical to the one used for the different primers pairs tested.

The Taq polymerase enzyme used was “Master Mix” (Fermentas GMBH, St-Leon, Rot, Germany) with a final primer concentration of 1.2 μM .

Ten microlitres of the amplified products were analysed by electrophoresis in 1% (w/v) agarose gels stained with ethidium bromide (0.5 mg mL^{-1}). The gels were visualised with an Ultraviolet Illuminator and Digitally recorder (GelDOC Bio-Rad, Hercule, USA). PCR was scored positive when a band of the appropriate size was visualised. The size of DNA fragments were estimated using the O⁺ Gene Ruler DNA ladder Mix (Fermentas GMBH, St-Leon, Rot, Germany).

Cloning and sequencing conditions

The desired bands after PCR amplification were cut from the gel and purified with QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The PCR products were cloned into pGEM-T Easy Vector (Promega Corp.), and the plasmids were introduced into *E. coli* JM109 cells by heat shock, following the manufacturer's protocol. Transformants were selected on LB agar medium complemented with ampicillin, isopropyl- β -D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside at a con-

centration of 100, 200 and 200 $\mu\text{g mL}^{-1}$, respectively. White colonies were picked and grown 24 h at 37°C in LB medium complemented with ampicillin at final concentration of 100 $\mu\text{g mL}^{-1}$. Plasmids were isolated from the transformed cells using the QiaPrep Spin Miniprep kit (Qiagen, Germany).

Restriction analysis of the plasmids was performed using *Eco*RI (Fermentas, Germany) to check for inserts. Cloned products were sequenced using Universal primers pUC-M13-R/F in an ABI PRISM dye terminator cycle sequencing kit (Cogenics S.A, France; MWG Eurofins, Martinsried, Germany).

The determined sequences were compared with the GenBank databases using the basic local alignment search tools (BLAST) software provided online by the National Center for Biotechnology Information (USA).

Silico tools

The nucleotide sequences used in this paper, which were extracted for GenBank accession numbers, are surfactin synthetases (CAB12142.1, CAB12143.1 and CAB12145.1), mycosubtilin synthetases (AAF08795.1, AAF08796 and AAF08797), plispastatin synthetases (CAB13717.1, CAB13716.1, CAB13715.1, CAB13713.1 and CAB13714.1) and fengycin synthetases (AAB80955, AAB00093, CAA09819 and AAB80956.1).

Results

Comparative analysis of the NRPS lipopeptides operons

In order to design specific primers from a lipopeptide family, we compared the similarity rates of synthetases in each family from *B. subtilis* at amino acid and nucleic acid levels (Table 2). The comparative analysis of lipopeptides at protein level showed a strong homology comprised between 75% and 90% for all synthetases confirming their membership of a same lipopeptides family. The nucleic sequences compared with those of the known lipopeptides of the same family showed a strong difference from one family to the other: DNA similarity was on average 63% for lichenysin and surfactin, 76% for mycosubtilin and the iturinA and only 43% for fengycin and plipastatin. The percentage of similarity is in the same range for different ORFs of each operon in the three lipopeptide families except for *mycC/ituC* where the similarity was 69.7% compared to the levels of homology for other ORFs of the same lipopeptide family (80%). This can be explained by the positions 6 and 7 of the amino acids: a sequence Ser-Asn is found in the mycosubtilin, while in iturinA, these amino acids are inverted (Asn-Ser).

Table 2 Comparative analyses of non ribosomal lipopeptides synthetases of *Bacillus* strains using Needle (Global) software

Sequence Origins	Lipopeptides NRPS genes	Protein similarity (%)	DNA similarity (%)
Surfactin synthetases of <i>B. subtilis</i> 168 and lichenysin synthetases of <i>B. licheniformis</i> ATCC14580	<i>srfA-A/lchA</i>	77.2	64.7
	<i>srfA-B/lchB</i>	77.1	65.5
	<i>srfA-C/lchC</i>	75.1	62.9
	<i>srfA-D/lchD</i>	75.6	62.8
Fengycin synthetases of <i>B. subtilis</i> F29-3 and plipastatin synthetases of <i>B. subtilis</i> 168	<i>ppsA/fenC</i>	87.6	43.4
	<i>ppsB/fenD</i>	89.6	44.3
	<i>ppsC/fenE</i>	90.0	43.1
	<i>ppsD/fenA</i>	88.1	42.9
	<i>ppsE/fenB</i>	88.4	42.8
Mycosubtilin synthetases of <i>B. subtilis</i> 6633 and iturinA synthetases of <i>B. subtilis</i> RB14	<i>mycA/ituA</i>	88.7	80.7
	<i>mycB/ituB</i>	88.6	79.8
	<i>mycC/ituC</i>	79.0	69.7

For protein (Method: Emboss/Matrix: Blosun62/Gap Open: 10, Gap Extend 0.5) For DNA (Method: Emboss/Matrix: DNA full/Gap Open: 10, Gap Extend 0.5)

Because of the low similarity of the nucleic sequences insides the fengycin family, two types of primers were designed, whereas only one was retained for the other families.

PCR primer design and in silico tests

Nucleotide sequences encoding the A and T domains of the modules of different synthetases were separately aligned for surfactin, plipastatin, fengycin and mycosubtilin genes. PCR primers were designed from the highly conserved nucleic acids motifs. A domains were used to design forward primers and T domains for the reverse ones. For example, the forward primer from mycosubtilin synthetase was designed from the alignment of the adenylation domain sequences of *mycA*, *mycB* and *mycC*; the reverse one from the alignment of all thiolation domains of the same sequences as determined by Vector NTI Advance 10™ software (Fig. 2). The same approach was used for the design of surfactin, fengycin and plipastatin synthetases primers. To ensure the highest hybridization criterions and to be able to amplify all the desired fragments after the choice of the most conserved regions, we degenerated only when needed and by preference the 3' end of the primers. IUPAC DNA codes were used to substitute nucleic acids that were not identical in the alignment. Other criteria were to balance the melting temperature (T_M) of a primer pair and to obtain amplicons between 300 and 1,200 bp. The H_yC was determined, and all primer characteristics are presented in Table 3. To validate these primers, in silico amplification was carried out, using AmplifX 1.4.4 software, on the NRPS genes of the different *Bacillus* strains (see Electronic supplementary material, Table 1).

The size of each fragment varied for each primer pair: for As1-F/Ts2-R, 419, 422, 425, 431; for Aml-F/Tml-R, 416 and 419; for Af2-F/Tfl-R, 443, 452 and 455 bp; and for Ap1-F/Tp1-F, 563, 893, 929, 947, 959 and 1,152 bp. The primer position in each aligned sequence can explain the difference of length of amplified products. The sequences obtained in silico were searched against NCBI's non-redundant database by using the BLASTX or BLASTN algorithms. The four primer pairs led to the selection of NRPS genes, and each primer pair amplified its corresponding sequences with 100% of homology. They also detected genes from the same lipopeptides family as lichenysin or iturin A synthetase genes by As1-F/Ts2-R and Aml-F/Tml-R primer pairs (<100% homology). All results proved the possible detection of NRPS genes with these primers.

Screening of lipopeptide NRPS genes

To assess the range of PCR tests specificity, the four primer pairs were checked with different *Bacillus* species and two NRPS non-producing *Lactobacillus* (Table 1). As expected, no amplification was obtained with the *Lactobacillus* strains tested. All fragments obtained were cloned and sequenced as described in "Materials and methods", and the BLAST results obtained with the different sequenced fragments are presented in Table 4. For each DNA isolated from *B. subtilis*, the size of the bands was similar to the size obtained in silico when the amplifications were positive (Fig. 3 and Table 3). A different intensity was observed for each primer pair and for each strain. A high intensity could correspond to a high number of fragments with a similar

Table 3 Characteristics of specific primers developed for the detection of the lipopeptide biosynthetic genes and results of the in silico tests

Lipopeptides	Primer name	Primer sequences ^a 5' 3'	Length (bp)	HyC ^b	Position in alignment sequences (vector N1)	Amplified products size (amplification, ×1.4-4)	Operon/gene origins strains	GenBank accession numbers of synthetases
Fengycins	AF2-F	GAATAYMTCGGMCGMTGKA	20	34	1,138-1,157	443, 452 bp	<i>Bacillus subtilis</i> F29-3	AF023464
	TF1-R	GCTTTWADKGAATSBCCGCC	20	72	75-94	455 bp		BACFENB CAA09819 AF023465
Pipastatins	Ap1-F	AGMCAGCKSGCMASATCMCC	20	64	694-713	893, 959, 929 bp	<i>Bacillus subtilis</i> 168	CAB13717.1
	Tp1-R	GCKATWWTGAARRCCGGCGG	20	32	70-89			CAB13716.1 CAB13715.1 CAB13714.1
Surfactins	As1-F	CGCGGMTACCGVATYGAGC	19	12	1,182-1,200	419, 422, 425, 431 bp	<i>Bacillus subtilis</i> 168	D13262.1
	Ts2-R	ATBCCCTTTBTWDGAATGTCGCC	23	36	76-98			AF233756
Mycosubtilins	Am1-F	CAKCAAGTSAATAATYCGMGG	20	34	1,178-1,197	416, 419 bp	<i>Bacillus subtilis</i> ATCC 6633	AF184956
	Tm1-R	CCDASATCAAAARAADTTATC	20	36	57-76			

^a Using IUPAC DNA code: Y=C or T, M=A or C, K=G or T, W=A or T, D=G, A or T, S=G or C, B=G, T or C, R=A or G^b Coefficient of hybridization

size as determined by in silico tests; likewise, the number of bands obtained with the Ap1-F/Tp1-R primer pair is probably due to the different size of amplified products confirming results obtained by in silico amplification.

Primer pairs were first tested with *B. subtilis* 168, *B. licheniformis* ATCC 14580 and *B. subtilis* ATCC 6633, which are partially or completely sequenced and are known to contain NRPS genes encoding synthetases involved in the biosynthesis of the three families of lipopeptides. Secondly, *B. subtilis* ATCC 21332 and ATCC 9943, which are described as lipopeptide producers, were used and finally *B. thuringiensis* serovar *berliner* 1915 and *B. cereus* LMG 2098 for which it is unknown whether or not they are NRPS producers. Prbac1/Prbac2 primers were used as a positive control.

Detection of previously sequenced NRPS genes involved in lipopeptide biosynthesis

As expected, NRPS genes involved in lipopeptide biosynthesis were detected by the primer pairs in *B. subtilis* 168, ATCC 6633 and *B. licheniformis* ATCC 14580. Indeed the As1-F/Ts2-R primers amplify a fragment of expected size, 100% identical to the surfactin synthetase gene in *B. subtilis* 168 and also detected the lichenysin synthetase gene in *B. licheniformis* ATCC 14580 with a high homology (99% and *e* value equal to 0.00). The detection of surfactin and lichenysin synthetase gene by As1-F/Ts2-R primers in *Bacillus* strains can be interpreted by the high similarity between them and explained by the membership to the same family of these peptides (Ongena and Jacques 2008). Ap1-R/Tp1-R primers highlighted the plipastatin synthetase gene in *B. subtilis* 168 strain (99% of homology with *e* value equal to 0.00) as expected. The detection of the mycosubtilin gene as described by Duitman et al. (1999) has been carried out by Am1-F/Tm1-R primers in ATCC 6633 with 99% of homology and *e* value=0.00.

Detection of non-sequenced NRPS genes in lipopeptide-producing strains

As1-F/Ts2-R primers amplify fragments of expected size, similar to the surfactin synthetase gene in *B. subtilis* ATCC 6633 (96% of homology with *e* value=0.00), *B. subtilis* ATCC 21332 (99%, *e* value=0.00) and *B. subtilis* ATCC 9943 (97%, *e* value=0.00), which are all characterised as surfactin-producing strains (Kluge et al. 1988; Duitman et al. 1999). The presence of the lichenysin synthetase gene was also detected in *B. licheniformis* ATCC 14580 by the Ap1-F/Tp1-R primer pair with 99% of homology and *e* value=0.00.

The iturin synthetase gene was detected in *B. subtilis* ATCC 9943 with a homology of 97% (*e* value=0.00) by the

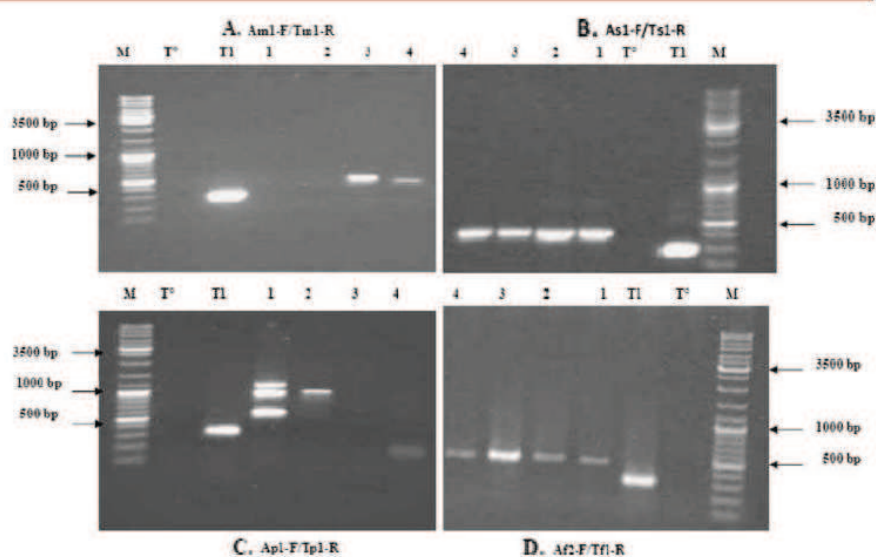
Table 4 Blast results of the sequenced products obtained from PCR amplification using gene-specific degenerated primers from biosynthetic genes of fengycin, pipastatin, surfactin and mycosubtilin

Isolates	Primer name	Product size (bp)	GenBank accession number of matching proteins obtained by Blastx	Match with probable proteins obtained by Blastx	Identity of corresponding gene by BlastN (%)	E value ^a of gene identity
<i>B. subtilis</i>	Δ2-F/Tp1-R	449	YP_03595850	Pipastatin synthetase of <i>B. subtilis</i> NCIB 3610	93	2e-176
<i>ATCC 21332</i>	Asl-F/Ts2-R	431	CAB12142	Surfactin A of <i>B. subtilis</i> 168	99	0.00
	Am1-F/Tm1-R	0				
<i>B. subtilis</i>	Ap1-F/Tp1-R	893	CAB13717	Pipastatin synthetase of <i>B. subtilis</i> 168	99	0.00
<i>ATCC 9943</i>	Δ2-F/Tp1-R	452	YP_001421438	Fengycin C of <i>B. amyloliquefaciens</i> FZ3-2	97	0.00
	Asl-F/Ts2-R	419	YP_001419996	Surfactin B of <i>B. amyloliquefaciens</i> FZB42	97	0.00
	Am1-F/Tm1-R	419	ABY89499	Iturin A synthetase B of <i>B. subtilis</i> MH25	97	0.00
	Ap1-F/Tp1-R	0				
<i>B. subtilis</i>	Δ2-F/Tp1-R	464	CAB13589	Polyketide synthetase type I of <i>B. subtilis</i> 168	88	3e-144
<i>ATCC 6633</i>	Asl-F/Ts2-R	419	CAA51223	Surfactin A of <i>B. subtilis</i> 168 <i>trpC2</i>	96	0.00
	Am1-F/Tm1-R	419	Q98900	Mycosubtilin synthetase B of <i>B. subtilis</i> 6633	99	0.00
	Ap1-F/Tp1-R	0				
<i>B. subtilis</i> 168	Δ2-F/Tp1-R	454	X172672.1	Operon of Surfactin of <i>B. subtilis</i> 168	99	0.00
	Asl-F/Ts2-R	425	CAB12142	Surfactin A of <i>B. subtilis</i> 168	100	0.00
	Am1-F/Tm1-R	0				
<i>B. licheniformis</i> ATCC 14580	Ap1-F/Tp1-R	959	CAB13717	Pipastatin A of <i>B. subtilis</i> 168	99	0.00
	Δ2-F/Tp1-R	532	YP_080976	Siderophore 2,3-DHB of <i>B. licheniformis</i> 14580	96	0.00
	Asl-F/Ts2-R	428	YP_077642	Lichenysin C of <i>B. licheniformis</i> 14580	99	0.00
	Am1-F/Tm1-R	0				
<i>B. thuringiensis berliner</i> 1915	Ap1-F/Tp1-R	951	YP_077640.1	Lichenysin A of <i>B. licheniformis</i> 14580	99	0.00
	Δ2-F/Tp1-R	0				
	Asl-F/Ts2-R	416	ACC55454	NRPS protein of <i>Acinetobacter baumannii</i>	88	8e-130
	Am1-F/Tm1-R	0				
<i>B. cereus</i> LMG 2098	Ap1-F/Tp1-R	0				
	Δ2-F/Tp1-R	0				
	Asl-F/Ts2-R	419	CAA51222	Surfactin ORF1 of <i>B. subtilis</i> MH25	91	8e-155
	Am1-F/Tm1-R	0				
	Ap1-F/Tp1-R	0				

^a E value (expect value) represents a measure of the probability of obtaining a result of an identity query better than the sequence shown.

For protein (Method: Emboss/Matrix: Blosum62/Gap Open: 10, Gap Extend 0.5) For DNA (Method: Emboss/Matrix: DNA full/Gap Open: 10, Gap Extend 0.5)

Fig. 3 PCR amplification by degenerated primers pairs with *B. subtilis*. **a** Am1-F/Tm1-R, **b** As1-F/Tf1-R, **c** Ap1-F/Tp1-R, **d** Af2-F/Tf1-R. *M* O'Gene ruler, *T*⁰ sample without DNA, *T* DNA with Prbac primers, *1* DNA *B. subtilis* 168, *2* DNA *B. subtilis* ATCC 21332, DNA *B. subtilis* ATCC 6633, DNA *B. subtilis* ATCC 9943



Am1-F/Tm1-R pair in this strain, the Af2-F/Tf1-R primer pair enabled us to detect the fengycin synthetase gene with 97% of homology and *e* value equals to 0.00. These results are in agreement with the identification of *B. subtilis* ATCC 9943 as a coproducer of the three families of lipopeptide (Vater et al. 2002; Chollet-Imbert et al. 2009).

Detection of unexpected genes

Ap1-F/Tp1-R primers amplify a fragment similar (99%, *e* value=0.00) to the plipastatin synthetase gene in *B. subtilis* ATCC 21332, which has not yet been identified as a producer of plipastatin molecules. This result confirms the detection of a fengycin family peptide in the supernatant of *B. subtilis* ATCC 21332 by Gancel et al. (2009).

All the primers were also used on two strains (*B. thuringiensis* serovar *berliner* 1915 and *B. cereus* LMG 2098) for which no NRPS gene or product was known. Only the As1-F/Ts2-R primer pair led to the detection of fragments similar to a “non-ribosomal peptide synthetase protein” gene in *B. thuringiensis berliner* 1915 (88% of similarity with an *e* value of 8e–130) and to a surfactin synthetase gene in *B. cereus* LMG 2098 (similarity of 91% with an *e* value 8e–155). These NRPS molecules could thus be new ones.

In addition, Af2-F/Tf1-R primers can detect several NRPS genes other than those of the family for which they are defined. Indeed, with this pair, surfactin synthetase gene (99% homology) and a siderophore 2,3-dihydroxybenzoate (96% homology) could be detected in *B. subtilis* 168 and in *B. licheniformis* ATCC 14580, respectively. This would suggest that the absence of fengycin family genes in a strain could lead to the detection of NRPS genes other than those for which primers were defined. Moreover, the Af2-F/Tf1-R

primer pair amplified in *B. subtilis* ATCC 6633, a fragment homologous to a polyketide synthase type I of *B. amyloliquefaciens* FZB42 (88%) which is similar to bacillaene synthetase of *B. subtilis* 168 (Chen et al. 2006, 2009). Recently, Chen et al. (2009) confirmed the hybrid NRPS/PKS structure of *bae* gene cluster, which encodes the bacillaene synthetase. This hybrid complex includes two NRPS modules (suggesting the presence of two amino acid building blocks). The Af2-F/Tf1-R primer would detect the NRPS portion of this hybrid.

Discussion

This approach, using degenerated primers designed using the alignment of adenylation or thiolation domains that compose lipopeptide synthetases, led to the detection of the presence of NRPS genes as confirmed for all amplified sequence fragments. When the result was positive, the sequencing of PCR products confirmed the presence of lipopeptide NRPS genes or other NRPS genes in the *Bacillus* genus. To summarise, in *B. subtilis* strains, the sequences of the PCR products showed a very high homology to the sequence of surfactin synthetase family genes when the primers As1-F/Ts2-R were used. Whereas the presence of surfactin molecules were described by many authors within *B. subtilis* species, it was surprising to detect surfactin gene in *B. cereus* LMG 2098. Until now, the only known lipopeptide molecule in *B. cereus* was the plipastatin (Umezawa et al. 1986). However, recently, Xiong et al. (2008) noticed in *B. cereus* Q1 (most similar to *B. cereus* LMG 2098) the presence of a *dhb* operon, encoding the biosynthetic template for the catecholic

siderophore, and a *sfp* gene, encoding a phosphopantetheinyl transferase required for the production of surfactin. They also detected the *mbtH* gene involved in mycobactin synthesis and a gene (BC2300) with unknown function between the *dhb* operon and the *sfp* gene. Whereas they did not find a surfactin synthetase gene (*srfAA*, *srfAB* or *srfAC*), they speculated that the three genes (*mbtH*, BC2300 and *sfp*) are located downstream of the *dhbF* gene and might belong to the *dhb* operon that is involved in antibiotics/siderophore/surfactin biosynthesis, a speculation that is wrong, as *dhb* was never involved in surfactin biosynthesis.

However, our results, by the detection of a fragment similar to surfactin synthetase in *B. cereus* ILMG 2008, could reinforce the hypothesis of the presence of the *SrfA* operon in some *B. cereus* strains.

Use of the *Apl1-F/Tp1-R* and *Aml1-F/Tm1-R* primers led to the amplification of plipastatins and iturins synthetase family genes, respectively, whereas *Af2-F/Tf1-R* primers detected fengycin, surfactin and polyketide synthetase genes. The observation that the *Af2-F/Tf1-R* Hyc is greater than 50 confirmed that these primers are more degenerated than the other primer pairs, and this could explain their lower specificity compared to the others. In addition, the amplification of the surfactin synthetase gene instead of plipastatin synthetase gene by *Af2-F/Tf1-R* in *B. subtilis* 168 was also confirmed by in silico analysis.

Overall, in other *Bacillus* species, primers amplify lipopeptide NRPS synthetase genes, such as the lichenysin synthetase gene in *B. licheniformis* ATCC 14580, or other NRPS genes, such as an amino acid adenylation synthetase in *B. thuringiensis* serovar *berliner* 1915. These results suggest that primers will preferentially detect the NRPS synthetase genes for which they are defined and, secondly, that, when these genes are not present, other genes containing adenylation and thiolation domains are detected as observed by primers from other species.

To date, to our knowledge, this is the first report that confirms the presence of the NRPS lipopeptides synthetase genes in the *B. subtilis* strains listed in Table 1, except for *B. subtilis* 168, which was completely sequenced by Kunst et al. 1997. This paper also confirms the fengycin production potential recently detected in our laboratory for *B. subtilis* ATCC 21332. Our PCR approach also highlighted the probable presence of polyketide synthase in *B. subtilis* ATCC 6633 similar to a Pks type I synthase in *B. subtilis* 168.

In conclusion, the design of degenerated primers on conserved nucleic sequence motifs after alignment of the adenylation and thiolation domains could be used to screen non-ribosomal synthetases genes in *Bacillus* spp. and could lead to the detection of new NRPS molecules. This method could also prove to be highly useful in multiplex PCR and in metagenomic screening.

Finally, our previous knowledge of the genetic potential of *Bacillus* strains for the synthesis of non-ribosomal metabolites could be greatly enhanced by the design of molecular tools for the rapid detection of NRPS genes by PCR.

Acknowledgements This work was supported by the Université des Sciences et Technologies de Lille, the Region Nord Pas de Calais, the Ministère de la Recherche Scientifique (ANR) and the European Funds for Regional Development. Arthur Tapi has a fellowship from Ivorian government. The authors thank William Everett for the re-reading of the manuscript.

References

- Ayuso-Sacido A, Genilloud O (2005) New PCR primers for screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microbial Ecol* 49:10–24
- Barrios-Llerena ME, Burja AM, Wright PC (2007) Genetic analysis of polyketide synthase and peptide synthetase genes in cyanobacteria as a mining tool for secondary metabolites. *J Ind Microbiol Biotechnol* 34:443–56
- Bonmatin JM, Laprevote O, Peypoux F (2003) Diversity among microbial cyclic lipopeptides: iturins and surfactins. Activity-structure relationships to design new bioactive agents. *Comb Chem High Throughput Screen* 6:541–556
- Caboche S, Papi M, Leclerc V, Fontaine A, Jacques Ph, Kucheryov G (2008) Norine: a database of non-ribosomal peptides. *Nucl Acids Res* 36:326–331
- Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsis A, Hiltzerth G, Grammel N, Stittmayer AW, Gottschalk G, Süßmuth RD, Borriss R (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol* 188:4024–4036
- Chen XH, Koumoutsis A, Scholz R, Schneider K, Vater J, Süßmuth R, Piel J, Borriss R (2009) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *J Biotechnol* 140:27–37
- Chollet-Imbert M, Gancel F, Slomianny C, Jacques Ph (2009) Differentiated pellicle organization and lipopeptide production in standing culture of *Bacillus subtilis* strains. *Arch Microbiol* 191:63–71
- Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC 6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc Natl Acad Sci USA* 96:13294–13299
- Eshita SM, Roberto NH, Beale JM, Mamiya BM, Workman RF (1995) Bacillomycin L₂, a new antibiotic of the iturin group: Isolations, structures, and antifungal activities of the congeners. *J Antibiotics* 48:1240–1247
- Gancel F, Monastrie L, Liu T, Zhao L, Nikov I (2009) Lipopeptide overproduction by cell immobilization on iron-enriched light polymer particles. *Process Biochem* 44:975–978
- Giacomodonato MN, Pettinari J, Souto GI, Meandez SB, Lopez NI (2001) A PCR-based method for the screening of bacterial strains with antifungal activity in suppressive soybean rhizosphere. *World J Microbiol Biotechnol* 17:51–55
- Grangemard I, Wellach J, Maget-Dena R, Peypoux F (2001) Lichenysin: a more efficient cation chelator than surfactin. *Appl Biochem Biotechnol* 90:199–210

- Jacques Ph, Hbid C, Destain J, Razafindralambo H, Paquot M, Pauw E, Thonart P (1999) Optimization of biosurfactant lipopeptide production from *Bacillus subtilis* S499 by Plackett-Burman design. *Appl Biochem Biotechnol* 77:223–233
- Jullien N (2007) AmpliX software 1.44. <http://ifrrj.nord.univ-mrs.fr/AmpliX-Home-page>. Accessed 12 Nov 2006
- Kluge B, Vater J, Salnikow K, Eckart K (1988) Studies on the biosynthesis of surfactin, a lipopeptide antibiotic from *Bacillus subtilis* ATCC 21332. *FEBS Lett* 231:107–110
- Kunst F et al (1997) The complete genome sequence of the Gram-positive model organism *Bacillus subtilis* (strain 168). *Nature* 390:249–256
- Lee YK, Kim SB, Park CS, Kim JG, Oh HM, Yoon BD, Kim HS (2005) Chromosomal integration of *sfp* gene in *Bacillus subtilis* to enhance bioavailability of hydrophobic liquids. *Appl Microbiol Biotechnol* 67:789–794
- Leclère V, Béchet M, Adam A, Guez JS, Wathélet B, Ongena M, Thonart P, Gancel F, Chollet-Imbert M, Jacques Ph (2005) Mycosubtilin overproduction by *Bacillus subtilis* BBG100 enhances the organism's antagonistic and biocontrol activities. *Appl Environ Microbiol* 71:4577–4584
- Marahiel M (1997) Protein templates for the biosynthesis of peptide antibiotics. *Chem Biol* 4:561–567
- Molenaar D, Bringel F, Schuren FH, De Vos WM, Siezen RJ, Kleerebezem M (2005) Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J Bacteriol* 187:6119–6127
- Moyne AL, Cleveland TE, Tuzun S (2004) Molecular characterization and analysis of the operon encoding the antifungal lipopeptide bacillomycin D. *FEMS Microbiol Lett* 234:43–49
- Nakano MM, Marahiel MA, Zuber P (1988) Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *B. subtilis*. *J Bacteriol* 170:5662–5668
- Needleman SB, Wunsch CD (1970) Needleman-Wunsch global alignment. *J Mol Biol* 48:443–453
- Neilan B, Dittmann E, Rouhiainen L, Bass RA, Schaub V, Sivonen K, Börner T (1999) Nonribosomal peptide synthesis and toxicity of Cyanobacteria. *J Bacteriol* 181:4089–4097
- Nikolskaya AN, Panaccione DG, Walton JD (1995) Identification of peptide synthetase-encoding genes from filamentous fungi producing host-selective phytotoxins or analogs. *Gene* 165:207–211
- Ongena M, Jacques P (2008) *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends Microbiol* 16:116–125
- Ongena M, Jacques P, Toure Y, Destain J, Jabrane A, Thonart P (2005) Involvement of fengycin-type lipopeptides in the multifaceted biocontrol potential of *Bacillus subtilis*. *Appl Microbiol Biotechnol* 69:29–38
- Peypoux F, Bonmatin JM, Wallach J (1999) Recent trends in the biochemistry of surfactin. *Appl Microbiol Biotechnol* 51:553–563
- Rajendran N (1999) Identification and cloning of a gene locus encoding peptide synthetase of *Pseudomonas fluorescens* by two sets of PCR primers. *Z Naturforsch* 54:105–109
- Ramarathnam R, Bo S, Chen Y, Dilantha WG, Xueven G, Kievit T (2007) Molecular and biochemical detection of fengycin and bacillomycin D-producing *Bacillus spp.*, antagonistic to fungal pathogens of canola and wheat. *Can J Microbiol* 53:901–911
- Rupf S, Merte K, Eschrich K (1999) Quantification of bacteria in oral samples by competitive polymerase chain reaction. *J Dent Res* 78:850–856
- Schneider J, Taraz K, Budzikiewicz H, Deleu M, Thonart P, Jacques P (1999) The structure of two fengycins from *Bacillus subtilis* S499. *Z Naturforsch* 54c:859–866
- Schwarzer D, Finking R, Marahiel MA (2003) Non ribosomal peptides: from genes to products. *Nat Prod Rep* 20:275–287
- Sieber SA, Marahiel MA (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* 105:715–738
- Steller S, Vollenbroich D, Leenders F, Stein T, Conrad B, Hofmeister J, Jacques Ph, Thonart P, Vater J (1999) Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chem Biol* 6:31–41
- Steller S, Sokoll A, Wilde C, Bernhard F, Franke P, Vater J (2004) Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* 43:11331–11343
- Turgay K, Marahiel MA (1994) A general approach for identifying and cloning of peptide synthetase genes. *Peptide Res* 7:238–241
- Tsuge K, Akiyama T, Shoda M (2001) Cloning, sequencing and characterization of iturinA operon. *J Bacteriol* 183:6265–6273
- Tsuge K, Matsui K, Itaya M (2007) Production of the non-ribosomal peptide plipastatin in *Bacillus subtilis* regulated by three relevant gene blocks assembled in a single movable DNA segment. *J Biotechnol* 129:592–603
- Umezawa H, Aoyagi T, Nishikiori T, Okuyama A, Yamagishi Y, Hamada M, Takeuchi T (1986) Plipastatins: new inhibitors of phospholipase A₂, produced by *Bacillus cereus* BMG302-ff67. I. Taxonomy, production, isolation and preliminary characterization. *J Antibiot* 39:737–744
- Vanittanakom N, Loeffler W, Koch U, Jung G (1986) Fengycin—a novel antifungal lipopeptide antibiotic produced by *Bacillus subtilis* F-29-3. *J Antibiot* 39:888–901
- Vater J, Kablitz B, Wilde C, Franke P, Mehta N, Cameotra SS (2002) Matrix-assisted laser desorption/ionization-time of flight mass spectrometry of lipopeptide biosurfactants in whole cells and culture filtrates of *Bacillus subtilis* C-1 isolated from petroleum sludge. *Appl Environ Microbiol* 68:6210–6219
- Wu S, Zhong J, Huan L (2006) Genetics of subpeptin JM4-A and subpeptin JM4-B production by *Bacillus subtilis* JM4. *Biochem Biophys Res Comm* 344:1147–1154
- Xiong Z, Jiang Y, Qi D, Lu H, Yang F, Yang J, Chen L, Sun L, Xu X, Xue Y, Zhu Y, Jin Q (2008) Complete genome sequence of the extremophilic *Bacillus cereus* strain Q1 with industrial applications. *J Bacteriol* 191:1120–1121
- Yakimov MM, Kroger A, Slepak TN, Giuliano L, Timmius KN, Golyshin PN (1998) A putative lichenysin A synthetase operon in *Bacillus licheniformis*: initial characterization. *Biochim Biophys Acta* 1399:141–153

Annexe 2. Extrait Publication 2

Detection of surfactins producing strains in Soumbala and Bikalga fermented condiments using Polymerase Chain Reaction and Matrix Assisted Laser Desorption/Ionization Mass Spectrometry methods

A. Savadogo^{a,b}, A. Tapi^b, M. Choller^b, B. Watheler^c, P. Jacques^b

^aCRSBAN, UFR-SVT-Université de Ouagadougou, Burkina Faso, BP 7021.

^bLaboratoire des Procédés Biologiques, Génie Enzymatique et Microbien, UPRES-EA 1026, Polytech'Lille/IUTA, Université des Sciences et Technologies de Lille, Av. Paul Langevin, 59655 Villeneuve d'Ascq cedex, France.

^cUnité de Chimie Biologique Industrielle, Faculté Universitaire des Sciences Agronomiques, B-5030 Gembloux, Belgique.

^{*}Correspondance to : Professeur Philippe JACQUES, Laboratoire des Procédés Biologiques, Génie Enzymatique et Microbien, UPRES-EA 1026, Polytech'Lille/IUTA, Université des Sciences et Technologies de Lille, Av. Paul Langevin, 59655 Villeneuve d'Ascq cedex, France

Tel : 0033328767440, Fax 0033328767401 (e-mail: philippe.jacques@polytech-lille.fr)

A. Savadogo and A. Tapi have contributed equally to the scientific work.

Keywords: Soumbala, Bikalga, *Bacillus*, non ribosomal lipopeptides, degenerated primers, surfactins.

ABSTRACT

Bacillus spp. are the predominant microorganisms in fermented African locust bean called Soumbala and Bikalga. The combination of PCR with non ribosomal specific primers and Matrix-assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) to identify non ribosomal peptides in these *Bacillus* strains which isolated from exotic foods has been reported.

Annexes 3 : Séquence Kurstakines

>readseq.input(1), 8634 bases, A9B30F8A checksum.

MTNEECYRLSHPQRRRIWYTEVIHSDKGICNLRFLFKLHRKMNYSKLNQVV
NRVISENDGLCIRLKENGHQEPKQYFIKHEEQEFEFDFSNVENTNLLEE
WIEQKTKDTFTFLNSDLYYFALIKLNDNECAVFGNVHHIIMDGLSIQIFT
EQIAKYYYEMHNENEKEIKNSYVQFLTNEQTYLNSSRFQKDQKFWLEEF
KTLPSVTGLKPYNTYQVSTKASRETFILTEHLKRKIEKFSEDYKISLYTF
FVAASFSLMYRWTSDDLIGLMAYGNRMTKMERKLI GMMVSTVPLRMDID
PEEEVLSFISRVSRKQSKLLRHQKYPFDLLKQINKENNSNVRLF GTTID
YRDRDESGDSLWIPNHEEVQDFAVHIENLIDIDSLHVHIDYREELFSKEE
IKRFFNVMLAIMENTFEKPKQKVAEIEIISEEDKRISLVEFNNPKLNVPP
QVTIHEIFEQQAMIYPNTIAVTYEQEKITYKELNERANQLAHYLQKKGVG
PDTLVGICVERSLEMIVGILGVLKAGGAYVPLDPTYPEQRLQYILEDASI
QLFVTQESLKEKWL PENVESICLDRDQDEIGKESKTLPLSNVGPKNLAY
VIYTSGSTGNPKGVMIEHNNVIRLFKSTDCWYQFNEKDTWTLFHSYAFDF
SVWEIWGALLYGGKLVVVPYWISRSKDFYQLLVEEEVTVLNQTPSAFRG
LIRVCEQEDKNKNLQLRYVIFGGEALEPISLLPWFQRYGEKNPQLINMY
ITETTVHVITYYPITLDDVQHASRSNIGKRIPDLEVYILDAYQQVPVIGVD
GELYIGGAGLARGYLNRPELTAERFISHPFSSDPKARLYRTGDLARYLPD
GNLDYRGRIDYQVKIRGFRIEIGIEIESTLHTYVYVKEAVVIVREDQPGDK
RLVAYVVDGDNVDGWREYLKAKLP SYMVP SGFVAMEAIPLTANGKVDREA
LPMPEEKQINSECVGPRNSNEQILTTIWKRVLGKVKVGIYDNFFEIGGDS
ILSIQIISQASQVGLKLT PKQMFECPTIAELAQVAIETQGVQAEQGIVIG
DVPLTPICQWFFEQEHPHTEHWNQSMLLRVKERLDVKLLEGAILNLLKHH
DALRFRYEQLPNGTWRQRNEETDELSVLHVVKRNQVNEKEWNTIIQEEMN
INQASFNLVTGPLMRVVYFEDNLGTNDRIFWVHHHLVVDGVSWRILLEDL
QVVYNQMKQGQEVRLPAKSTSFKEWSERLQTYSDSGISKEVQNYWNQOVE
KETMKIPMDYPIQETTESIDQVTRTLGIEETHELLEHVPVTHKTRIDEV
LLTALGQAI VDC TNQQT VSIHLEGHGREVIEGIDL SRTVGWFTSIYPVH
LNFQGTQTPIEGLKSVEQLRRIPNRGV DYGILRYLNKGLLPFYQQKPSI
SFNYLGQFDQVFSRESLFMQETGFTFLDHAPDSKPSHLIDVIGMVKDEKL
HFVWMYSREQFSKLKIQLIADGMLRHLRQLINKPTTESAFTVSDFAADAED
LTEESLSKVLLKLKKKRVMGDVIDIYGLSPLQKGILFHTLYDQDDHSHFS
YIVQVGFLLGGQLDIATFEKAWYVIHRHEVLRSVFVWEEVEKPLQAVYQ
RLPFTIHQEDWSTYPHDERAKKLQQFLNVDRRKGFLLDEAPLMRVTAIKE
AKEETRIIWTHHHILLDGWVSLSVFSVMEVYLMKIKGELLNLKPPLPYK
KYIQWINKQDQSQAEEFWKEELKGFYAPTPLAMERKEQGQEKGYGECAYQ
VSEELTESLQEWVRNSRLTLNTLVQGAWAYLMSRYSGSDVVFVGT SAGR
PTDLMGAENMVGLFINTLPTRIRLTDNTSVVDWLKRIQMKEMERRQYEYN
SLIDIQGWSEVPRNSSLFHTLYVFENYPIQGERNDLRLLDVTS TEQNTY
PLTLIVMPGKKITLKLMYDRSYFNEETINLIQNHLQLIILQMTKSLDSKL
FEITHTEEEQIQLLEEWNHTQVNYSAEGMIHTMFEEQVKKTPEAIAASY
ENEQITYKELEKRANQLAHYLQKHGVPESLVGVYMERSLQMMIALLGIL
KSGAAYVPLDPTYPESRLRYILEDAGIEVLVTEENSKNLFVSENIE TICM
NKDYTAIEKEESTPCISGVGTENLAYVMYTSGSTGNPKGVMIEHHSVINY
LEWMQHKYPLSEKDVVLQKTPFSFVSVWELFWGIHV GASVSFLPPGGEK
DPSIIAEVIKQHQTIVQFVPSMLS VFLDHFNRIELNMNCSSVRHVFSGG
EELSSGLVRRFQKWNYSQGVKLTNFYGPTEATYVNAFDIQPNQEFVSI
GQPIQNTQLYVLDQNQRLQSIGIEGELYIGGAGLARGYLNRPNLTAEKVF
SHPFQGERLERTGDSVRYLTDGNIEFIGRMDHQVKMRGFRIELGEIEAT
LEKCSFIEKAVVLVREDRPGDQRLVAYVISDENTEWEVREYLQKQLPNHMI
PAHFVELEHFP LTPNGKIDRKALPIPDEQAIGDLNVLPRTPSEELIASVW
GQVLGIENIGIQDSFFERGGHSLLATQIVSRLQELFQIKIPLRELFKHD
VEALAKRVDQLRKEDKEREVAPLVPMKREEV IPLSYAQKRLWFIDQLMPN
SAMYNIHAACRLTGKWSIEALEVGWNQLLERHESLRTTILEQEDEFPQQV
QSHVFRHIPQTNLTDLSLEDREMERKRLIQNEAEFPFHLGGQPLIRTRIL
KVDREEWVLICTMHIIISDGWSMGILLEEWMAFYEGALTGKPVELKELSI
QYADFVMWQKEWQKEENLNQHLQYWKELSGELPVLQLPMDRPRPAVQTH
RGASQSLIVANSIQEKLKDL SLQEGCTLFMTLMAAYQSFLSRYTGQDDII
VGSPIANRNVKIEGLIGFFANTLVYRADFSENPTFQELLSQVRKKALKKA
YEYQDIPFEKVVESVKPERNTSHSPIFQTMFTLQNTKQEFQLSERQMEL
MESHASVAKFDLSLFASTEELSLTFEYNTDLFNDTTIERMASHFKHWL
NQIVSHPKCVLSELNMLSKVEYKQLLEEWNKSRVIDMEESTIHTMFEEQV
KKTPEAIAVVCDEELTYRELDERSNQLAHYLQKNGVTCESLVGVCVTRS
SEMI VGLLGIMKAGGAYVIDPAYPESRLQYILED AQIKVLVTQEKLQQK
MVIPKFVNVICIDRNRAETIEQEVTTVCTSEVTGDNLAYIITYTSGSTGNPK
GVMIEHRNTVTMIHWAHTYSRKELAGVLASTSLSFDLSVFEVFPVPLTMG
GKVI V TENALHLEKLSKGVTLVNTVPSAAKELVRAKTIPSSSVKVMNLAG
EPLPYSLVQDLYERNITIEKVYNYG PSED TTYSTYMELEKGMHRVPPIG
KPIFNTEVYVLSAEQKMPIGVVGELYIGGSGLARGYLKRPDLTKNRFIP
HPFKEGERIYRTGDLVRYLPDGNLEYVGRIDHQVKIRGFRIELGEIEAIL
QRHTLVNEAIVMVREDYPGDPRLIAYVVG DGAQKWRDYLKDQLPNYMIIP

AYFVGLNAFPLTPNGKIDRKALPAPEKQEIASGYIAPRTSTTEKTIVSIWH
EVLGMENIGIQDSFFEIGGHSLLATQAVSNLKEAFAIDIPLHDLFMFHTV
EQLAEQIDQLLDNKNQEMQNGSKENVSLQDYIQKEAEVSNDEELLELLKQ
LEELSEEEAQGIFESNLIIEGVKKMRSDLLAKLSSLSPEKRAWLKQMQK
KENKEALPLSYAQQLWFMDRFPNPNSSLYNIPTVWHLKGWNIPEALEKGF
NRLIERHESLRTVFKEVGEQPVQQIVEFLPRALPVRDYSQLPLEVKEKEV
DSLIAREAQEPFNLNMNGPLVRNQLVQLEKDEWLLCTMHIIISDAWSIGI
FMNELLAFYEEETGGNPAKLSSLSIQYADFAKWQKEWLQGDVLRNQLTYW
QEELSGELPILQLPVDRPRPVKQTYSGATHHVIFFPKLLSQLKDISRQEG
STLFMTLMAAYQSFLARYTGQKDILVGSPIANRNHKGVEGLIGFFVNTLV
YRSDLGSIPTFREILNQTKKKALKAYEYQDIPFEKVVEAVQPERSTSHSP
IFQTMFTLQNIKQERLDPNRSIEMVESNMSIAKFDLSLTAYEVEEGLFV
SFYNTDLFDSSTIARMAGHFENWLNETIYHPDESYSKLSMLSDTEQKQL
LEEWNDTDVVYGHDCMIHELFEQQVARTPNAVAVVYEGGKLTQYQELNEKS
NQLAHFLQKRGIGPESLVGICIERSPDMIIGLFGILKAGGAYVPLDPSYP
ENRLRYILENSQIQVLLTKEALQDWLPKDIRAICLDRDQVMISKESNLAP
VSGVTANNLAYIITYTSGSTGNPKGVMIHHSVINRLQWQKQYPLSEDDT
ILQKTPFSFVSVWELFWWSFVGARVCLLPPEGKEKDPVIEEYIERYRVS
TMHFVPSMLSTFLDYMELYNKRQDVSSLIQVFTSGEALNTEQVRRFKGIF
YNVQQTKLINLYGPTTEATVDVTTYDCDLEKEAMLPIGRPIDNTELYVLD
QYQVVPVIGVAGELYLGVGGLARGYFNRPDLTTERFIHPFKDGERLYRT
GDLVRYMNDNRNLEYIGRIDNQVKIRGFRIELGEIEAALHDHSSVKEAVVL
VKEDRPGDKLVAYVVGEQDTEWREYLLKQLPHYMVPAYFFQIEGMPLT
PNGKVNRRKALLEEQFISEDITSSRTPVEELIVSVWSQVLGIKNI SVQD
SFFEIGGHSLLATQVVSRLQEIFQIELPVRELFYTTVESLAKRLEQLRK
GDKKREIPLIPMERGEAIPLSYAQQLWFIDQFTPNALYNMMPVCRLT
GNWLLALEETGWNQLIERHESLRTVFQEVNGQPVQQIKSYAFQSIKPTDL
TMLSSDREEEVKRFIQQETEVPPDLTEGPIRTSILHVGEEEWILLCTL
HHIISDGWSMGILLEEWMAFYEKATDGKVAELEPLPVQYADFAQWQKQWL
KEEVLDDQLQYWREELSSELPLVQLPMDRPRPAIQTHHGSTYTLVLPSTL
HDKLNELSRKEGATLFMTLLAAYQSFLSRYTGQEDILVGSPIANRNRYREI
EGLIGFFVNTLVYRTNLSGGPTFQDVLYQVRQKALKAYEYQDIPFEKIVE
VVQPERSTSHSPIFQTMFILQNMKQEFVLSRSSRIEMIESHSPIAKFDLS
VMAVETEGLLFTFEYNKDLFNATTIERMAGHFEEKWLYEVSHRPQNPLHD
LSMLSEPERNLLLETWNDTVMEMSHQGLICERFEEQVAMPDAIAIVDQT
KQWYTGELDIQANQLANVLQRKGVAPESVVGVLPRSAELMASLLGILKA
GGAYVVLDPYPKSRLEYMIEDAGIQFVVTAEVQEGHFAHVEMVRLEELT
VESVIAPTRQINPENLAYIVYTSGSTGKPKGVMEYRS LMNMVSWHQUEVY
RISAEDRATQIAGVAFDSAVQEIWPLYTAGAALYLSTEELRINPEALRDW
LIDSRITASFAPTPILERLLKLSWPKTDLRFIITGGDQLTQYPSNDIPF
AVINQYGPSENTVVTDCYVPVGVTTGTPTSIGRPIANTEVYVLDLHQLV
PIGVIGDLYIGGKSLARGYANRPDLTKEKFIHPFKSGERLYYTGDKASY
LSDGNLQFHGRIDDQVKIRGFRIELGEIEAVLQAHSVKEAVVLVREDNQ
GDKRLVAVVVGEGSVHEWREHLQTHLPNYMVPNTNFIEMEFLPLTPNGKLD
LKALPILSEQSAENTVYPRTPLEELIASVWSQVLGIRDIDVQDSFFELGG
HSLLATQVVSRLQEVFQIKLPVRELFYESTVEALSKRIDQLRKSGSEKREI
PPLMPIERGGNIPLSYAQQLWFIDQFAPNSALYNMPPMACRLTGNWLP
LELGNWQLIERHESLRTVFYEEDGHPIQQIQPYVFRPLQMDLTKLSLEE
REKLEQWIQTEVESPFDEQGPLFRGKLIRISEEWEILLCNMHIIISDG
WSMEILLQEWMAFYENAI GEKQAELEPLPVQYADFAQWQREWLKDELLYQ
QLAYWKEELSGELPVLMLPMDRPRPVQTHHGLTHNVLLSRSLDDKLNEL
SRQEGATLFMTLLAAYQSFLARYTGQTDIIVGSPIANRNRYREIEGLIGF
VNTLVYRADLSNAPAFQELLSQVRIKALKAYEYQDVPFEKIVEVIQPERN
TSHSPIFQTMFTLQNTGQKLPELHGRNIEVMESNAPIAKFDLSLTAAEVE
EGLLLTFVYRTDLFDSLTI ESMAEHFGNWLNVIVENPDKSLAKLPILSGL
QQKQLEEWNNDAVAYLQESTIHQLFEEQVNRTPDAVAVVDEKQQLTYREL
NEKANQLAHYLQCCGIGNESLVGLCFERSVEMVVSIMGIWKAGAAVPLD
PSYPESRLRYILEDTGIIQVLVTNESLEDWIPKEMKIVCLDRDQAMISQES
ILSPKCEVTGETLAYVIYTSGSTGNPKGVLIQHHSVLNLSHGLQKEVFEH
EIPSNMHVGLNASTAFDASIQQQLQMLLYGSSLYIIPNEVRSDPEQFVAYI
RENKLEIFDITPSLLQLLIDVGLLETCDGIHVPSKVLVGGEAIMPSLWEQ
LVETDKIQFYNVYGPTECTVDATCYHIKKDSKRVITIGRPLPNVQTYVLD
NRLPVPVGVGMELYIGGAGLARGYLNRPELTSERFISHPFKEGERLYRTG
DLVRYLPDGNIDYLRMDNQVKIRGFRIELGEIESTLQEHDLVKEVVIV
RENQSGDKRLVAVVVGESLEEWREYLLTKLPIHMI PSYFVEMKELPLTI
NGKVDRKSLPIPDYQGTQEGYAAPRNEREHKLSIIWERVLSKRIGIYDN
FFEIGGDSILSIQIVSRAKQAGLQTPKQLFEHQTI AELAQVVKEEQGVQ
AEQGIITGETILTPIQQRFFAQNHPPNPHHWNQSMFFRTKEKLDIGSLEKA
ARSLLLHHDALRLRYERLPNGTWEQWNEGIEQLMLTVISLDEVPEADWH
EVIQKEINVAQERLNLHDGPLMRMVYFDEDEKTGRLFWTIHHLAVDGVSW
RILLEDLTQAYTQAVQGGKIQLPKMKSTSFKEWSEKLQHYAETGISKEVLH
YWEQQAQEDVILPVDGTISNPVSAVTEETVVLNENETRMLLQETLSH
RVQINEVLLAALVQATAACTGQPILSVGLEGHGREIIEDVDLSRTVGWF
TSIYPVHLNITNANTPIAALKAVKEQVRKIPNKGVDYGVLCYMNATMCEQ
LSFQYTPSISFNYLQGFQDLFSSDAMFIPENEFKRLDHAAGSKRSHVIDV
IGVVTDGKLQFTWYVNVGQFAKSTIQSIAQNMLYQLSTLIQFSRESALT
ISDFAMANLSQEGLTNVLNKMHRGKNNQITDLYPLSPQEGMIFHTLHDQ



GDEHVAPYIVQLSFMIRGKMNIPTFEQAWKSVIQRHEIFRTAFVWDEIEE
PVQVVYENIPFKVNKEDWRTMTSEEIEEKRKVFLELDRKQAFQFDEAPLM
RVTVIQEGEEYRIVWTHHHILLDGWSLPLVFNELLTVYQKRMNGEAVKL
PKSSPYKKYIQWLREQDKEQAEQFWREKLKGFTAPTLLGLESKEEEKGYT
EKVTYLSEEQTQALQGWAQRNKLTLSTVIQGAWAYLMSRYSGENDIVFGV
TSSGRSTEIIDVENIVGPFITTSPTRIQLMDDIKLIDWLQKIQEEIERR
QYFYASLTEIQAWSEVPRGIPLFNSLYVFENYPVKEESSENLEIGELEGV
EETHYPLGLAVVPGIQLSLKLMYDRSKFNRVTIERMLGHLHKVLMQMLKN
IDQNLSELVYITEAEQRKLLLEEWNNNTISYPRENAIHQLFEEQVNRTPD
LAVVDEKQQLTYRELNERANQLAHLQKCGIGTESLVGLCFDRSDEMVG
LMGIWKAGAAVPLDPSNPESRLRYILEDTDIQVLVTNEVLVGWIPKEIK
VVCLDRDQAMISRESKLSPTCEVTGENLAYVIYTSGSTGNPKGALVQHHS
VINLSYGLQKEVFSHEIPNNMRVGLNASIAFDVSIQQLQMLLYGSSLYII
PSEVRTNPEQFVSYIRENKLEMFDMTPSLLQLLIDGGLLEISADVHVPSK
ILVGGEAIMPSLWEQLVETEKIDFYNVYGPTACTVDATCYHIKKDSKRV
IGRPLPNVQTYVLDKNLLPVPVGVGTGELYIGGAGIARGYLNRPELTAERF
IRHPFNEEERLYRTGDLVRYLPDGNLDYLRRIDNQVKIRGFRIELEIEA
NLERHPLVKEAVVLVTEDKLGEQRLVAYVVGDSMHDWREYLKTQVPNYM
VPAHFIVKDEIPLTTNGKVDRKALNNLTILRENNFSTPIIPRDEIEYQLI
KIWQDVLQIEDVYVNDFFNRGGHSLLVIKLVSKIREKFGKEIKVSTLIK
NPTVEGIACLIRDKHDMTKSVAALVPLQESEKRPFFCVHPFMGNVFCYIQ
LARLLKSRCSFYGLQNPLIEKKEIDELTLPEVIQLYMBEIKRVQPEGPYR
LGGWSLGGTIAYEIATVLKNQGEVEELLVMDTKVPSPQDHKTSDEMLSY
IYEHFIPRDLTEQEGDLVHKQDMLVEQLIMEGVLTDPADLINLKQVINAH
RKCLNIMAEDLIPYSGEVLYFSAEEGKELFTDWGPLLKGRVNRYSVPGS
HEEIVDSPAVEKIAMYLLNEFEGIKEKSILLTEK

Annexes 4 :

Code génétique

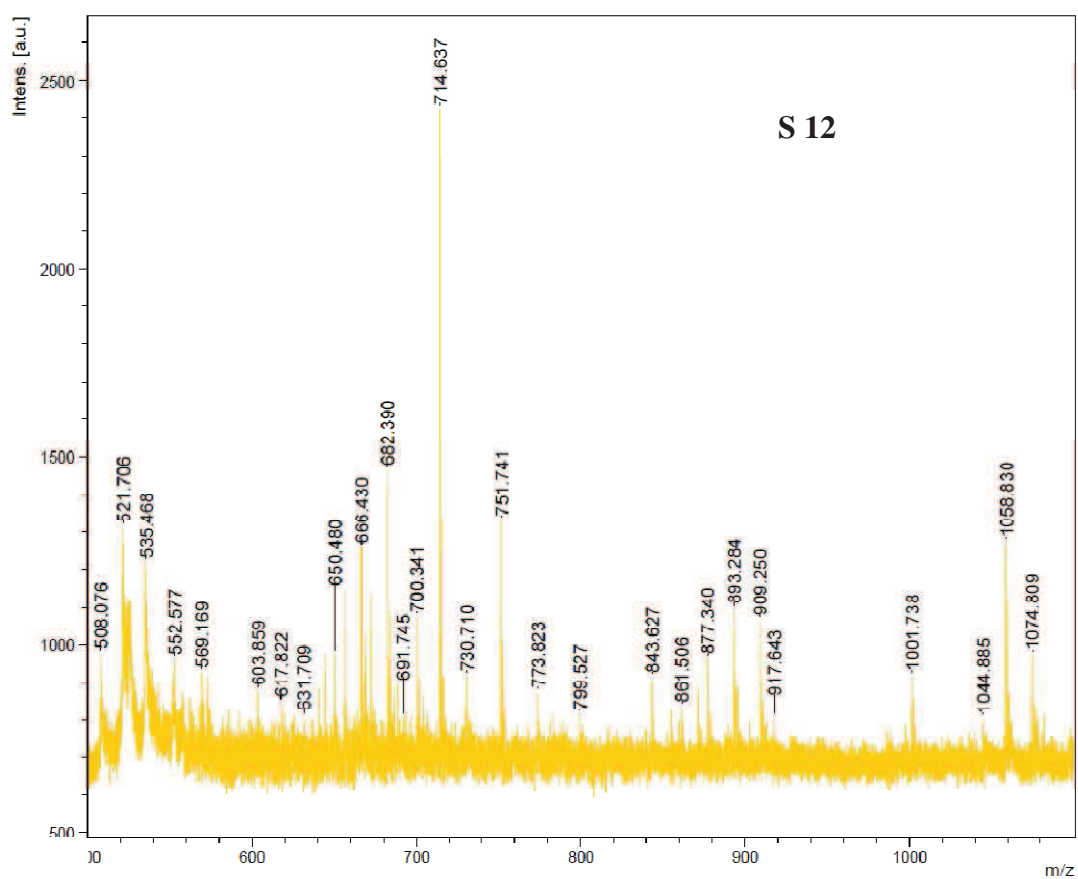
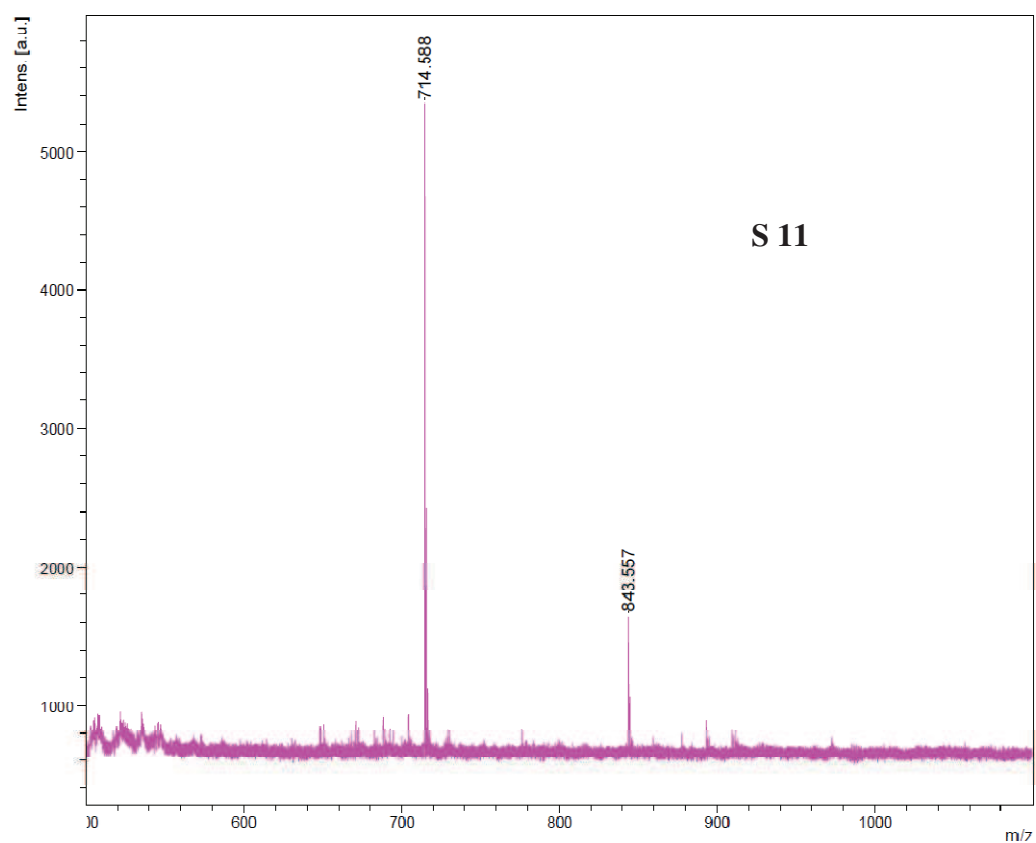
Dans le tableau suivant, la rangée indique la première lettre du codon, la première colonne indique la seconde et la dernière colonne indique la dernière lettre du codon.

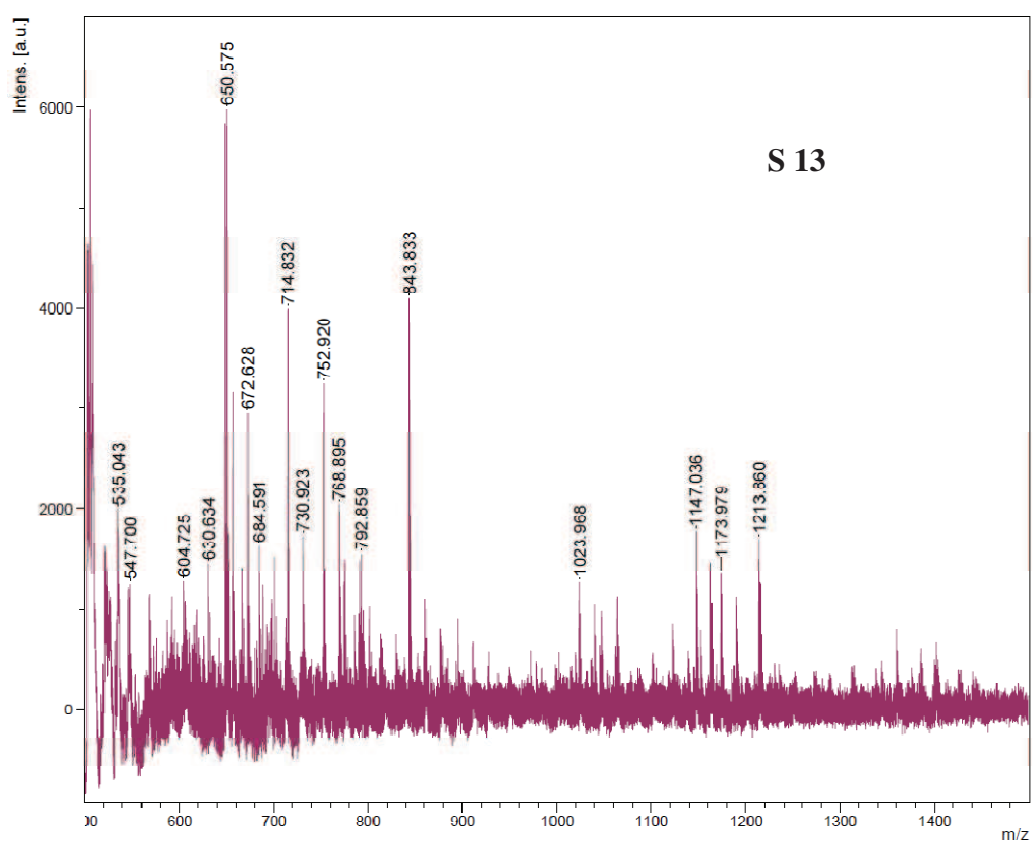
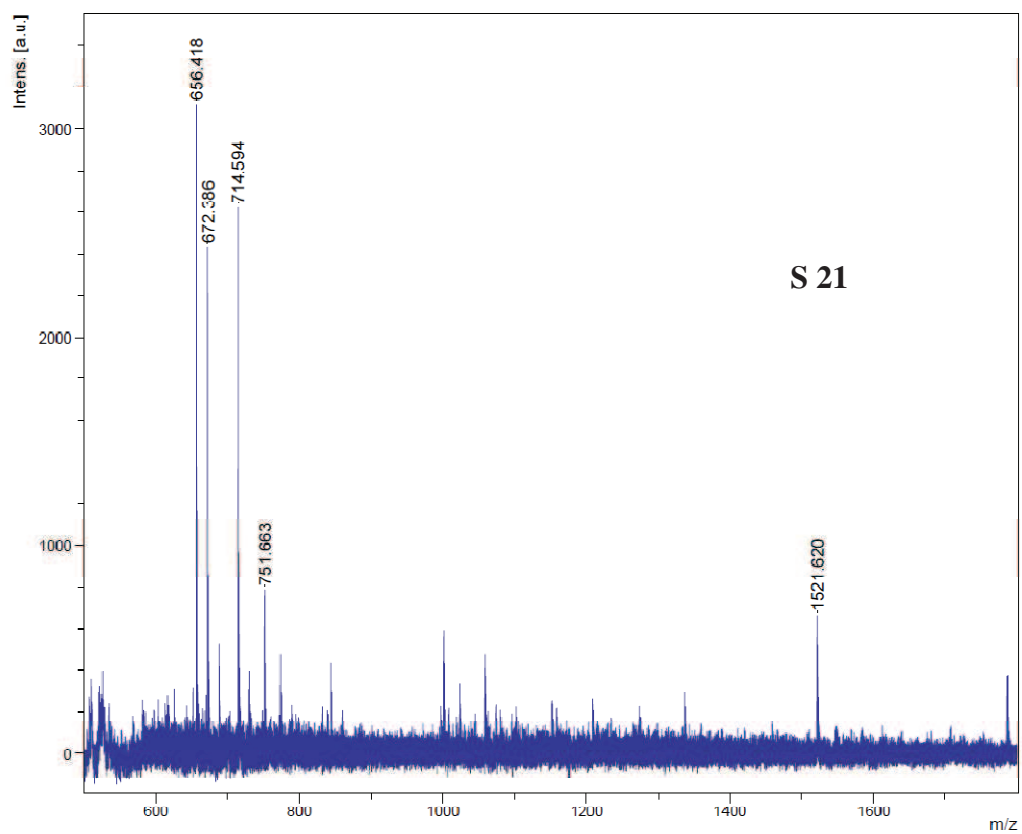
	U		C		A		G		
U	UUU	phénylalanine	UCU	sérine	UAU	tyrosine	UGU	cystéine	U
	UUC		UCC		UAC		UGC		C
	UUA	leucine	UCA		UAA	stop	UGA	stop/sélenocystéine	A
	UUG		UCG		UAG		UGG	tryptophane	G
C	CUU	leucine	CCU	proline	CAU	histidine	CGU	arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	isoleucine	ACU	thrénine	AAU	asparagine	AGU	sérine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lysine	AGA	arginine	A
	AUG	méthionine/start	ACG		AAG		AGG		G
G	GUU	valine	GCU	alanine	GAU	acide aspartique	GGU	glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	acide glutamique	GGA		A
	GUG		GCG		GAG		GGG		G

Acides aminés et masse molaire (g/mol)

Acides aminés	Code	MW (g/mol)
Alanine	Ala	89
Arginine	Arg	174
Asparagine	Asn	132
Aspartate	Asp	133
Cystéine	Cys	121
Glutamate	Glu	147
Glutamine	Gln	146
Glycine	Gly	75
Histidine	His	155
Isoleucine	Ile	131
Leucine	Leu	131
Lysine	Lys	146
Méthionine	Met	149
Ornithine	Orn	132
Phénylalanine	Phe	165
Proline	Pro	115
Sérine	Ser	105
Thrénine	Thr	119
Tryptophane	Trp	204
Tyrosine	Tyr	181

Annexe 5





Résumé

Les synthétases non ribosomiales sont des complexes multienzymatiques qui synthétisent des peptides atypiques par un mécanisme de thiotemplate indépendant des ribosomes. Ces peptides possèdent un large spectre d'activités biologiques et sont utilisés dans les domaines de la médecine et de la recherche biologique. Leur application dans d'autres domaines comme l'agriculture et l'environnement est de plus en plus envisagée à cause de leurs fonctions diverses et surtout de leur innocuité par rapport aux analogues chimiques. Plusieurs microorganismes possèderaient la capacité de synthétiser ces molécules particulières.

Nos travaux se sont concentrés sur la mise au point d'une nouvelle approche moléculaire de criblage par la réaction de polymérisation en chaîne, dans le but de détecter des souches de *Lactobacillus* et de *Bacillus* capables de produire ces peptides particuliers. Ainsi, plusieurs couples d'amorces dégénérées spécifiques (As1-F/Ts2-R, Am1-F/Tm1-R, Af2-F/Tf1-R, Ap1-F/Tp1-R et Abl1-F/Tbl1-R) ont été définis à partir des gènes connus de synthétases impliquées dans la biosynthèse de ces molécules. Ces amorces ont permis de révéler le potentiel de synthèse de molécules non ribosomiales des souches de *Bacillus cereus* LMG 2098 et de *Bacillus thuringiensis* serovar *berliner*, puis ont confirmé la présence de gènes NRPS dans d'autres bacilles déjà connus pour la synthèse de ces molécules.

La combinaison de ce criblage moléculaire avec les analyses biochimiques classiques a permis de montrer que la kurstakine présente chez *B. thuringiensis* est synthétisée par voie non ribosomiale.

Mots clés : Lipopeptides non ribosomiaux, amorces dégénérées, PCR, kurstakines.

Abstract

Non-ribosomal peptide synthetases are multimodular enzymes that make atypical peptides through a thiotemplate mechanism independent of ribosomes.

These peptides have a broad spectrum of biological activities and are used in the fields of medicine and the biological research. They also have a lot of potential applications in the agricultural and environmental fields because of their various functions and especially their harmlessness compared to the chemical analogues. Several bacteria would have the capacity to synthesize these particular molecules.

Our works concentrated on the development of a new molecular approach to screen by PCR *Lactobacillus* and *Bacillus* strains able to produce these particular peptides. Thus, several specific degenerated primer pairs (As1-F/Ts2-R, Am1-F/Tm1-R, Af2-F/Tf1-R, Ap1-F/Tp1-R and Abl1-F/Tbl1-R) were defined starting from known genes of synthetases implied in the biosynthesis of these molecules. These primers reveal the potential of synthesis of non ribosomal molecules by *Bacillus cereus* LMG 2098 and *Bacillus thuringiensis* strains, and then confirmed the presence of NRPS genes in other *Bacillus* already known for the synthesis of these molecules.

The combination of this molecular screening tool with the traditional biochemical analyses made it possible to show that the kurstakin presents in *B. thuringiensis* is synthesized by non ribosomal pathway.

Keywords: Non ribosomal lipopeptides, degenerated primers, PCR, Kurstakins.