

---

Université de Lille 1 Sciences et Technologies

# Thèse

Présentée par Thibault CARADEC

Pour l'obtention du titre de

Docteur de l'Université de Lille 1 Sciences et Technologies

Ecole doctorale Sciences de la Matière, du Rayonnement et de l'Environnement

Spécialité : Ingénierie des fonctions biologiques

---

## **Etude du mécanisme de production de deux nouvelles molécules synthétisées par la voie non-ribosomique**

Préparée au laboratoire ProBioGEM

**CONFIDENTIEL**

soutenue le 13 mai 2015

Philippe Jacques, Professeur (ProBioGEM, Université Lille 1)	Directeur
Bertrand Aigle, Professeur (Université de Lorraine)	Rapporteur
Tilmann Weber, Senior Resarcher (Technical University of Denmark)	Rapporteur
Didier Lereclus, Directeur de recherche (INRA MICALIS UMR1319)	Examineur
Valérie Leclère, Maître de conférences (ProBioGEM Université Lille 1)	Co-directrice
Marlène Chollet, Maître de conférences (ProBioGEM Université Lille 1)	Encadrante

## Remerciements

Je remercie le Professeur Pascal DHULSTER de m'avoir donné l'occasion de réaliser les travaux présentés dans ce manuscrit au sein du laboratoire ProBioGEM de l'Institut Charles Violette.

Je remercie mon directeur de thèse le Professeur Philippe JACQUES pour avoir assuré la direction de ces travaux et la confiance qu'il m'a accordé pour la réalisation de ce projet.

Je remercie mes encadrantes Valérie LECLERE et Marlène CHOLLET pour leur encadrement rigoureux, leurs nombreux conseils et leur aide précieuse pour la réalisation de ces travaux et la rédaction de ce manuscrit

Je remercie vivement les membres du jury Bertrand AIGLE, Tilmann WEBER et Didier LERECLUS d'avoir accepté d'évaluer ces travaux.

Je remercie particulièrement Corinne BOISTEL pour l'aide fournie pour la réalisation des travaux sur *Bacillus thuringiensis*, pour sa rigueur et son efficacité, et pour son agréable compagnie. Un grand merci à Gabrielle CHATAIGNE pour la réalisation des analyses de spectrométrie de masse, ainsi que pour les nombreux conseils sur les méthodes de purification et d'analyses des molécules étudiées lors de ces travaux. Merci à Max BECHET pour ses nombreux conseils pour la réalisation des travaux de biologie moléculaire.

Merci à François COUTTE pour ses conseils et son aide pour les premiers tests de production de kurstakine en fermenteur, pour ses conseils sur la purification et l'analyse des lipopeptides de *Bacillus*, et pour son aide à l'encadrement des étudiants dans le cadre de l'enseignement en GB-IAAL.

Un remerciement particulier va à toutes les personnes avec qui j'ai pu collaborer dans le cadre de ces travaux :

Merci au personnel de l'équipe MICALIS de l'INRA de Jouy en Josas pour ces travaux conjoints et pour leur expertise de *Bacillus thuringiensis*, en particulier Didier Lereclus, Michel Gohar et Sébastien Gélis.

Merci aux personnes du groupe de travail du projet européen MESIAB pour les travaux communs sur *Penicillium islandicum* et *Aster tataricus*.

Merci à tous les stagiaires ayant participé à ce projet : Rémi, Naomi, Sidi, Axel, Camille et Aymeric.

Je remercie vivement les doctorants du laboratoire m'ayant permis de travailler dans une ambiance chaleureuse et détendue. Merci à ceux qui étaient là quand je suis arrivé, nouveau doctorant à ProBioGEM : Omar, Amirouche, Marwan, Aurélien, Layal, Adil. Merci à ceux qui sont arrivés après moi : Rémi, Kalim, Juliette, Qassim, Alaa, Yazen, Ameen, Ahmed, Debarun, Alexandra.

Merci à tout le personnel du laboratoire ProBioGEM pour les nombreux échanges qui ont nourri ce travail.

Merci à Cathy OUBLION pour sa réactivité et son aide précieuse.

Quelques remerciements particuliers pour Joani et Pauline avec qui j'ai passé d'excellents moments.

Merci à Ségolène et Amandine pour nos repas conviviaux et leur bonne humeur.

Merci à Cédric pour tous les bons moments passés au laboratoire, à nos nombreuses discussions autour de nos passions communes.

Merci à ma famille qui me soutient depuis le début de mes études et pour leurs encouragements régulier.

Merci à Nicolas pour son soutien au quotidien.

## Résumé

La voie de synthèse non-ribosomique représente une source importante de molécules d'intérêt. Mettant en jeu de larges complexes multi-enzymatiques, les Synthétases de Peptides Non Ribosomiques (NRPS), ce mode de production aboutit à la formation de peptides originaux d'une grande diversité structurale et fonctionnelle. Les problématiques de recherche de nouvelles molécules d'origines naturelles pour une utilisation dans différents domaines industriels font des molécules d'origine non-ribosomique de parfaites candidates pour la découverte des médicaments, surfactants ou antibiotiques des prochaines années.

Les travaux décrits ont pour but l'étude du mécanisme de production de deux familles de molécules issues de la synthèse non-ribosomique, la kurstakine, un lipopeptide produit par *Bacillus thuringiensis*, et les PPCC (PentaPeptides Cycliques Chlorés), regroupant plusieurs molécules d'origine eucaryote. L'étude de la production de kurstakine par une série de mutants de *B. thuringiensis* a permis d'identifier plusieurs facteurs limitant la production de cette molécule, et d'ouvrir la voie à la mise en place de procédé de production en milieu liquide de cette molécules. L'étude des PPCC a permis d'identifier plusieurs gènes NRPS potentiellement impliqués dans la production de ces familles de molécules dans le génome de deux moisissures.

## Abstract

The Non Ribosomal Synthesis pathway is a very important source of molecules of interest. Based on large multi-enzymatic complexes, the Non Ribosomal Peptide Synthetases (NRPS), this production mode leads to the formation of original peptides, with a large structural and activities diversity. The need of new molecules from natural sources, turns the non-ribosomal molecules into a promising source of drugs, surfactant or antibiotics for the future.

This work aim to the understanding of the production mechanism of two families of non-ribosomal molecules, the kurstakin, a lipopeptide produced by *Bacillus thuringiensis*, and the CCPPs (Chlorinated Cyclic Penta Peptides), regrouping several eukaryotic molecules. The study of the kurstakin production using a series of *B. thuringiensis* mutants led to the identification of several limiting factors for the production of this molecule, and opened the possibility to the setting up of production process in liquid medium of this molecule. The study of CCPPs led to the identification of several genes, potentially implied on the production of these molecules in the genome of two fungal strains.

# Table des matières

Remerciements	2
Résumé	4
Liste des abréviations	8
Avant-propos	10
<b>CHAPITRE 1 INTRODUCTION</b>	<b>12</b>
1. Synthèse Bibliographique	13
1.1 La voie de synthèse non-ribosomique	13
1.1.1 Découverte	13
1.1.2 Structure des synthétases	13
1.1.3 Fonctionnement	14
1.1.4 Domaines enzymatiques des NRPS	14
1.1.5 Diversité des peptides issus de la voie non ribosomique	21
1.1.6 Outils bioinformatiques pour l'étude des NRPS	25
1.2 La kurstakine	30
1.2.1 Structure et découverte	30
1.2.2 Mécanisme de synthèse	31
1.2.3 Régulation	32
1.2.4 La kurstakine, un lipopeptide intracellulaire ?	33
1.2.5 La kurstakine, biomarqueur de <i>Bacillus cereus</i>	33
1.2.6 Les différentes familles de lipopeptide produits par <i>Bacillus</i>	34
1.2.7 Activités biologiques des lipopeptides de <i>Bacillus</i>	36
1.2.8 <i>Bacillus thuringiensis</i>	37
1.3 Les PentaPeptides Cycliques Chlorés (PPCC)	40
1.3.1 Les astines	40
1.3.2 La cyclochlorotine et l'hydroxycyclochlorotine	43
1.3.3 Mécanisme de synthèse des PPCC : Un mécanisme NRPS ?	45
2. Situation et objectif des travaux	47
<b>CHAPITRE 2 Matériels et méthodes</b>	<b>49</b>
1. Souches et plasmides utilisés	50
1.1 Souches microbiennes	50
1.2 Plasmides	51
1.3. Milieux de culture	52
Milieu LB	52
Milieu 863	52
Milieu Czapek	53
Milieu AK agar #2	53

<i>Milieu PDA</i>	53
<i>Milieu SD medium-uracil</i>	53
2. <i>Production de kurstakine</i>	54
2.1 Production en milieu solide	54
2.2 Production en milieu liquide	54
2.2.1 <i>Production en fioles Erlenmeyer</i>	54
2.2.2 <i>Production en fermenteur</i>	55
2.3 Purification	55
2.4 Détection et quantification	56
3. <i>Production de PPCC</i>	56
4. <i>Techniques analytiques</i>	57
4.1 HPLC	57
4.2 Spectrométrie de masse	57
5. <i>Techniques de biologie moléculaire</i>	58
5.1 Extraction de l'ADN de moisissure	58
5.2 PCR	59
5.3 Digestions enzymatiques et Gel d'agarose	60
5.4 Clonage gap-repair cloning	61
6. <i>Tests d'activité biologique</i>	62
6.1 Souches cibles d'activité antibiotique	62
6.2 Test d'activité antifongique.	63
6.3 Test d'activité antibactérienne	63
7. <i>Analyse in silico</i>	64
7.1 Florine : Workflow d'identification de NRPs à partir de séquences génomiques	64
7.2 Outils généraux d'analyse bioinformatique	67
7.3 Outils spécifiques à l'analyse de NRPS	70
<b>CHAPITRE 3 Résultats</b>	<b>74</b>
1. <i>Etude du mécanisme de synthèse de la kurstakine</i>	75
1.1 Introduction	75
1.2 Etude du mécanisme de production de la kurstakine	76
1.2.1 <i>Détection de la kurstakine produite par Bt HD-73 et Bt407</i>	76
1.2.2 <i>Analyse quantitative de production de kurstakine par différentes souches de différentes souches de B. thuringiensis et B. cereus</i>	78
1.2.3 <i>Analyse bioinformatique de la kurstakine synthétase chez Bacillus thuringiensis HD-73</i>	82
1.2.4 <i>Cinétique de production de la kurstakine</i>	86
1.2.5 <i>Etude de la relation entre la sporulation de Bt et la production de kurstakine</i>	95

1.2.6 Effet de l'inactivation de la production de toxines de <i>B. thuringiensis</i> sur la production de kurstakine	96
1.2.7 Effet de l'augmentation de la stabilité de l'ARNm de la kurstakine synthétase sur la production de kurstakine	101
1.2.8 Rôle du gène <i>krsE</i> dans le transport de la kurstakine	103
1.3 Tests d'activités biologiques de la kurstakine	110
1.3.1 Test d'activité antifongique	110
1.3.2 Tests d'activité antibactérienne et antifongique en plaque (CMI)	112
Conclusion et perspectives	113
<b>2. Etude du mécanisme de production des PPCC</b>	<b>117</b>
<b>2.1 Introduction</b>	<b>117</b>
<b>2.2 Etude du mécanisme de production des PPCC</b>	<b>118</b>
2.2.1 : Confirmation de la production de peptides de <i>Penicillium islandicum</i>	118
2.2.2 Analyse bioinformatique du génome de <i>P. islandicum</i>	118
2.2.3 Identification d'une moisissure productrice d'astines.	131
2.2.4 Analyse bioinformatique du génome de l'endophyte d' <i>A. tataricus</i>	133
2.2.5 Analyse de l'activité des domaines A des synthétases de <i>P. islandicum</i> .	133
Conclusion et perspectives	139
<b>CHAPITRE 4 Conclusion</b>	<b>142</b>
Références Bibliographiques	149

## Liste des abréviations

μL : microlitre  
ACN : Acétonitrile  
ADN : Acide désoxyribonucléique  
AK : Arret Kirshbaum  
AMP : Adénosine Monophosphate  
ARNm : Acide ribonucléique messenger  
ATP : Adénosine Triphosphate  
BLAST : Basic local alignment search tool  
Bt : *Bacillus thuringiensis*  
CCT : Cyclochlorotine  
CMC : Concentration micellaire critique  
CMI : Concentration minimale inhibitrice  
Da : Dalton  
FPLC : Fast protein liquid chromatography  
G : accélération  
HCCT : Hydroxycyclochlorotine  
HPLC : Chromatographie Liquide Haute Performance  
LB : Luria Bertani  
M : Molaire (mole/L)  
MALDI-TOF : Matrix Assisted Laser Desorption Ionisation Time Of Flight  
mL : millilitre  
MS : Spectrométrie de masse  
MUSCLE : MUltiple Sequence Comparison by Log-Expectation  
NCBI : National Center for biotechnology information  
nm : nanomètre  
NRPS : Synthétase de Peptide Non Ribosomique  
PCP : Peptidyl Carrie Protein  
PCR : Réaction de polymérisation en chaine  
PDA : Potato dextrose agar  
PEG : Polyéthylène glycol  
PKS : Polykétide Synhtase  
Ppant : Phosphopantéthéinyl transférase  
PPCC : PentaPeptide Cyclique Chloré  
PPi : Anion Pyrophosphate

rpm : Rotation par minute  
SD medium : Synthetic dextrose medium  
Spe : Solid phase extraction  
SVM : Machine à vecteurs de support  
TBE : Tris Borate EDTA  
TFA : Acide trifluoroacétique  
V/V : volume pour volume  
WL : Weblogo

## **Avant-propos**

Dans de nombreux champs de recherche, la découverte de nouvelles molécules représente un défi important pour répondre aux problématiques de divers secteurs économiques et industriels, tels les domaines de la pharmacie, de l'agroalimentaire ou de l'industrie cosmétique. Les avancées dans le domaine de la synthèse chimique ont permis la découverte de nombreuses molécules répondant à ces problématiques en termes d'activité, de production, de rentabilité et de performance.

Néanmoins, la prise de conscience de l'impact de l'utilisation massive de molécules obtenues par voie chimique, notamment sur la santé humaine et l'environnement, souligne l'importance de la découverte de nouvelles molécules répondant aux mêmes critères d'efficacité ou économiques, mais présentant un impact moindre sur la santé et/ou l'environnement. Cette prise de conscience est notamment illustrée au niveau européen par la mise en place de la norme sur l'enregistrement, l'évaluation, l'autorisation et la restriction des produits chimiques REACH, (pour Registration, Evaluation, Authorization and Restriction of Chemical substances) faisant porter aux industries la responsabilité d'évaluer et de gérer les risques posés par les produits chimiques, afin d'assurer un niveau élevé de protection de la santé humaine et de l'environnement.

Ainsi, de nombreux efforts ont été réalisés ces dernières années pour la découverte, la production et l'optimisation de molécules naturelles. Celles-ci possèdent plusieurs avantages sur les molécules obtenues par synthèse chimique, notamment leur biodégradabilité et le renouvellement souvent plus simple de leur source d'obtention. De plus, les produits d'origine naturelle bénéficient généralement d'une meilleure image auprès du consommateur en comparaison avec les produits issus de la synthèse chimique.

Parmi les molécules d'origine naturelle, les peptides issus de la voie non-ribosomique possèdent une place particulière. Produits selon une voie de synthèse biologique originale, mettant en jeu des synthétases peptidiques non-ribosomiques (Non Ribosomal Peptide Synthetase : NRPS), ceux-ci présentent une grande variété de structures et d'activités, dues notamment à l'originalité du mode de fonctionnement des NRPS, et qui sont peu ou pas retrouvées chez des peptides produits à l'issue de la traduction d'ARN messagers.

Néanmoins, l'utilisation de peptides non-ribosomiques amène de nouvelles problématiques. Ainsi, les voies de régulation et les mécanismes de production de ces métabolites secondaires sont particulièrement complexes, et nécessitent donc une approche spécifique pour

l'identification et l'optimisation de la production de ceux-ci. Une étude des mécanismes de synthèse aboutissant à la production des molécules d'intérêts est nécessaire de façon préalable à la mise en place de procédés de production ainsi que l'étude du potentiel industriel de ces peptides non-ribosomiques.

Les travaux présentés dans ce manuscrit ont pour but l'étude du mécanisme de synthèse de ces peptides :

- La kurstakine, lipopeptide potentiellement antifongique produit par le groupe *cereus* des *Bacillus* et dont le mécanisme de production a été en partie identifié de façon préalable à ces travaux, avec l'identification des gènes de la kurstakine synthétase.
- L'astine, peptide utilisé dans la médecine traditionnelle chinoise comme antitussif et possédant une activité antitumorale, connu comme étant un produit de la plante *Aster tataricus*, et qui présente de très fortes similarités de structure avec la cyclochlorotine, peptide produit par la moisissure *Penicillium islandicum*.

L'étude de ces deux molécules a ainsi permis de caractériser les mécanismes de synthèse non-ribosomiques chez deux modèles, l'un procaryote, et l'autre eucaryote, et de mettre en évidence les différences entre ces deux modèles.

En premier lieu, une synthèse bibliographique présentée ici a pour but de réaliser un état de l'art des connaissances disponibles dans la littérature scientifique sur le mécanisme de synthèse non-ribosomique, sur les outils de bioinformatique pouvant être utilisés dans le cadre de l'étude de NRPS, mais également sur les mécanismes de production et les organismes producteurs des deux molécules étudiées.

Les différentes méthodes utilisées dans la réalisation de ces travaux seront ensuite décrites dans le chapitre matériel et méthodes.

Les résultats obtenus pour chacun des deux modèles étudiés seront ensuite présentés dans le troisième chapitre.

Enfin, une discussion globale des résultats obtenus ainsi qu'une conclusion sur les travaux réalisés et les différences liées à l'étude de deux modèles différents formeront un dernier chapitre.

# **CHAPITRE 1**

## **INTRODUCTION**

# 1. Synthèse Bibliographique

## 1.1 La voie de synthèse non-ribosomique

### 1.1.1 Découverte

La voie de synthèse peptidique non-ribosomique est une voie originale de production d'une grande variété de peptides, découverte à la fin des années 60. Cette voie de synthèse diffère de la voie « classique » de synthèse de protéines, ou traduction, présente chez tous les organismes vivants (Tomino *et al.*, 1967). En effet, la production de protéines de la voie classique suit les étapes suivantes :

-Les gènes contenus dans l'ADN de l'organisme sont transcrits en ARN messagers.

-Les ARNm sont traduits selon le code génétique en séquence d'acides aminés, formant alors la structure primaire des protéines.

Les protéines sont donc le produit de la traduction directe de l'information contenue dans l'ADN en acides aminés correspondant.

Dans le cadre de la synthèse non ribosomique, des complexes multienzymatiques, les synthétases peptidiques non ribosomiques (NRPS, de l'anglais Non Ribosomal Peptide Synthetase), sont à l'origine de la production de métabolites secondaires de nature peptidique chez les microorganismes (Gevers *et al.*, 1968 ; Weckermann *et al.*, 1988 ; Kleinkauf et Döhren, 1990).

### 1.1.2 Structure des synthétases

Ces synthétases, qui fonctionnent comme des chaînes d'assemblage de peptides, sont divisées en différents modules, composés chacun d'un ou plusieurs domaines enzymatiques (Weber et Marahiel 2001 ; Samel *et al.*, 2008), (Figure 1).

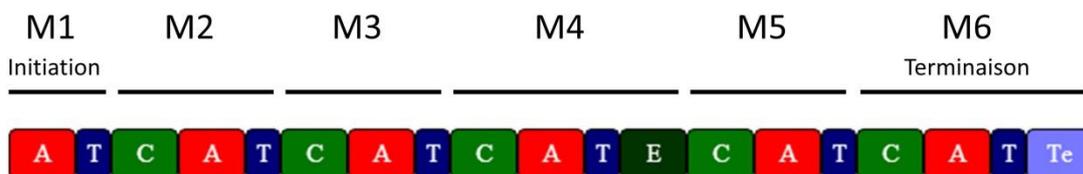


Figure 1 : Schéma de la structure globale d'une synthétase NRPS. Les modules, représentés par les barres noires sont numérotés de la partie N-terminale à la partie C-terminale. Chaque module contient au minimum un domaine C, un domaine A et un domaine T à l'exception du premier module M1, appelé module d'initiation, sauf pour certaines synthétases de lipopeptides. Le dernier module, appelé module de terminaison, contient généralement un domaine de thioestérase.

### 1.1.3 Fonctionnement

La structure modulaire des NRPS explique leur mode d'action spécifique. En effet, les synthétases vont agir comme des chaînes d'assemblage, chaque module effectuant la sélection d'un monomère devant être intégré dans la chaîne peptidique du produit final. Chaque module de la synthétase va effectuer une série de réactions enzymatiques au moyen de différents domaines enzymatiques (Figure 2) (Schwarzer *et al.*, 2003 ; Finking et Marahiel, 2004). Dans le cadre du mode de synthèse linéaire de peptide, l'intégration de monomère au peptide par chaque module va s'effectuer de l'extrémité N-terminale à l'extrémité C-terminale de la NRPS.

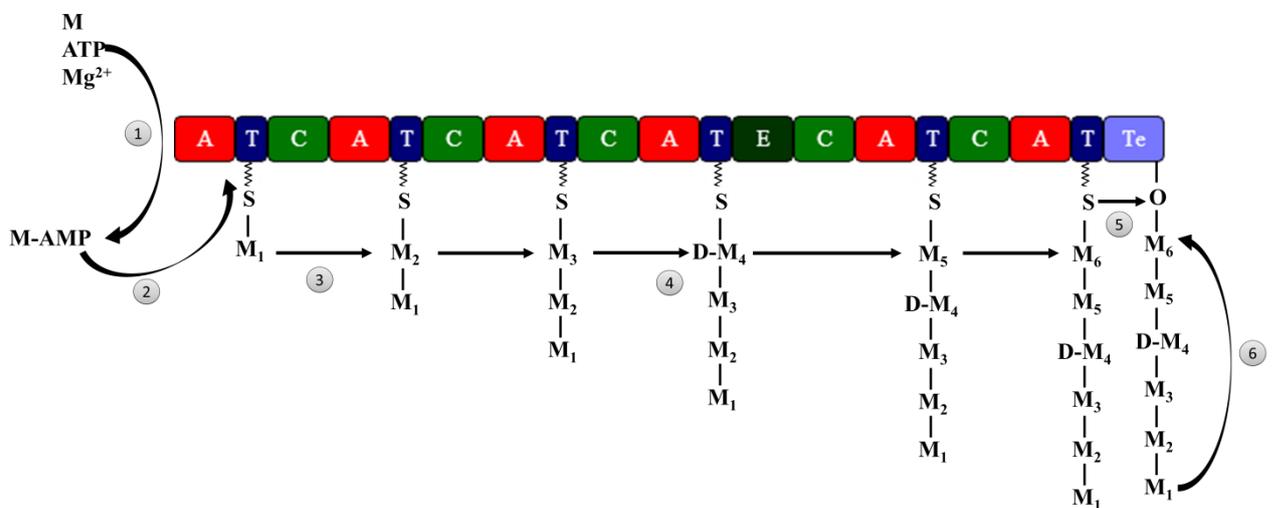


Figure 2 : Schéma de fonctionnement global d'une NRPS (A : Domaine d'adénylation ; T : Domaine de thiolation ; C : Domaine de condensation ; E : Domaine d'épimérisation ; Te : Thioestérase ; 1 : Activation d'un monomère en aminoacyl-AMP par le domaine A ; 2 : Transfert du monomère activé au domaine de thiolation ; 3 Condensation du monomère lié au domaine T ; 4 : Possibilité de modification d'acide aminé, par exemple une épimérisation ; 5 : Transestérification ; 6 : Relargage du peptide formé par hydrolyse ou cyclisation (D'après Strieker *et al.*, 2010)

### 1.1.4 Domaines enzymatiques des NRPS

Un module type d'une NRPS est composé de 3 domaines enzymatiques :

-Le domaine d'adénylation (domaine A), responsable de la reconnaissance plus ou moins spécifique d'un acide aminé et son activation en aminoacyl-adénylate en présence d'ATP.

-Le domaine de condensation (domaine C), responsable de la formation de la liaison peptidique entre le monomère porté par le domaine précédent, et celui porté par le module auquel ce domaine de condensation appartient.

-Le domaine de thiolation (domaine T, ou PCP pour Peptidyl Carrier Protein) lie le peptide en formation à la synthétase par l'intermédiaire d'une liaison thioester. Pour cela, un cofacteur de type 4'-phosphopantéthéinyl est requis, permettant la formation de la liaison covalente entre la synthétase et le monomère activé (Lambalot *et al.*, 1996).

#### 1.1.4.1 Domaine d'adénylation

La réaction catalysée par le domaine d'adénylation est représentée dans la Figure 3. Cette réaction nécessite la présence de  $Mg^{2+}$  et consomme une molécule d'ATP.

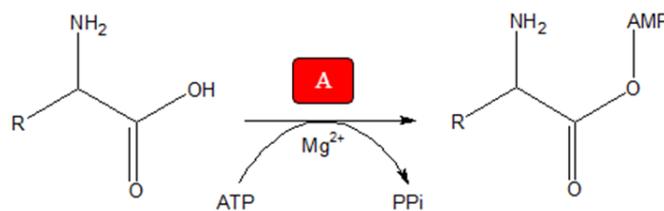
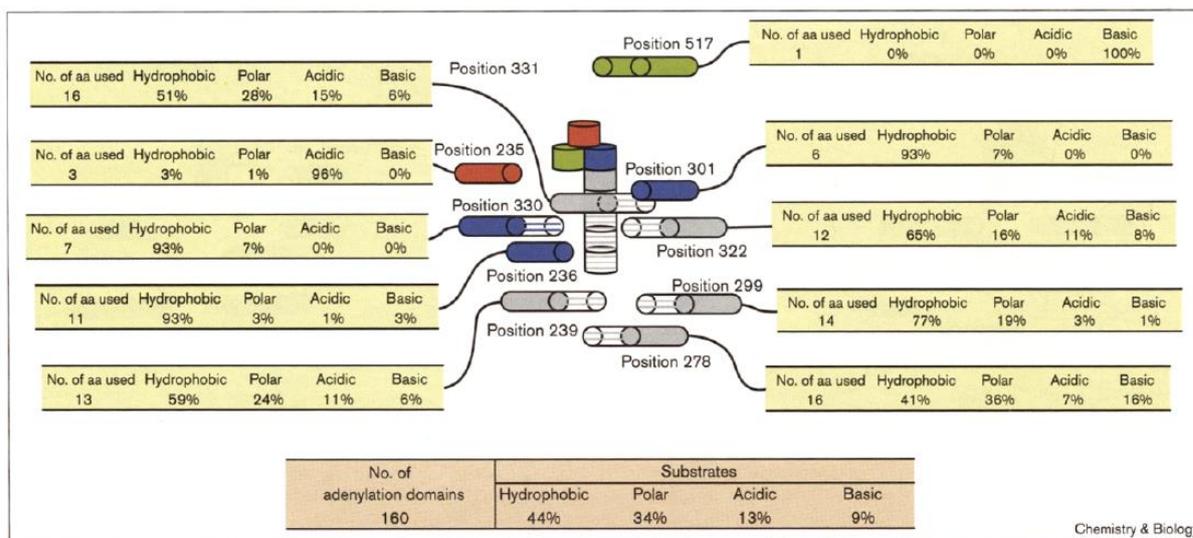


Figure 3 : Schéma de l'activité enzymatique du domaine d'adénylation.

Les domaines d'adénylation (domaine A) activent spécifiquement des monomères qui sont intégrés au peptide. La sélectivité des domaines A a été étudiée pour la première fois à la fin des années 90, et une meilleure compréhension des mécanismes mis en jeu dans le cadre de cette sélectivité a été obtenue par l'étude de la structure de la partie du domaine A interagissant avec le monomère incorporé. Ainsi l'élucidation de la structure tridimensionnelle par cristallographie d'un domaine A activant une phénylalanine, PheA, provenant de la gramicidine synthétase GrsA de *Bacillus brevis* a permis d'identifier la position de 10 acides aminés formant le site de fixation de la phénylalanine au sein de ce domaine. Par extension, une étude *in silico* de 160 domaines d'adénylation a permis de mettre en relation les acides aminés présents aux positions précédemment identifiées chez PheA et le substrat activé par ces différents domaines d'adénylation (Figure 4). Ainsi, le code de spécificité des domaines A a pu être identifié, et désigné « code de Stachelhaus », selon le nom de son découvreur, Torsten Stachelhaus (Stachelhaus *et al.*, 1999).



Chemistry & Biology

**Figure 4 : Schéma de la composition des acides aminés du site actif en interaction avec le substrat du domaine d'adénylation. Les positions indiquées correspondent à la position de chaque site dans la séquence du domaine d'adénylation (Stachelhaus *et al.*, 1999).**

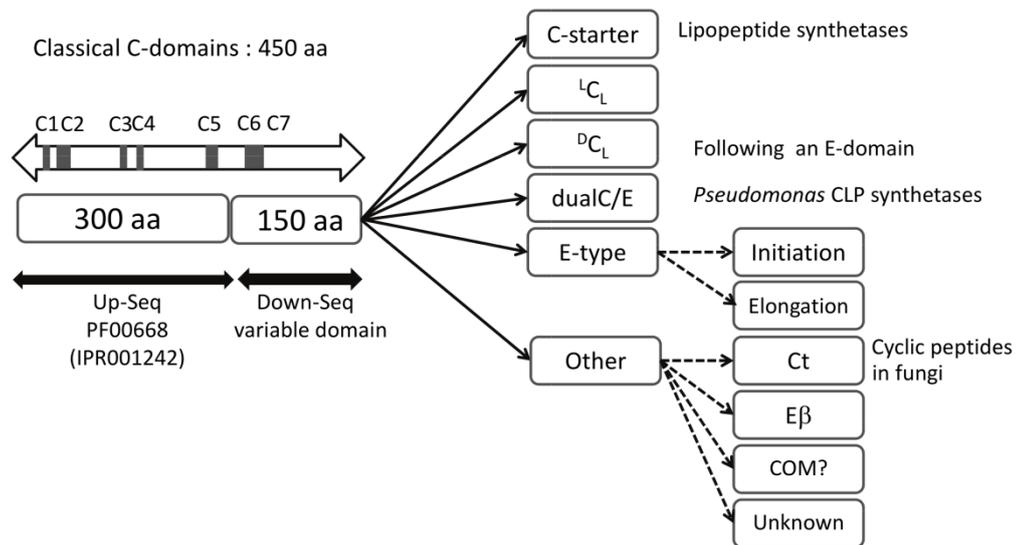
Une autre méthode de prédiction des monomères activés par les domaines A au moyen de techniques de machine à vecteurs de support (ou SVM, de l'anglais *Support Vector Machine*) a également démontré que les 34 acides aminés situés dans un rayon de 8 Å du site de fixation du substrat du domaine pouvaient également jouer un rôle dans la spécificité des domaines A (Rausch *et al.*, 2005). La comparaison de ces deux méthodes a démontré que si le code de Stachelhaus était très efficace pour la prédiction spécifique d'un monomère activé par un domaine A, le méthode mise au point par Rausch *et al.*, permettait d'obtenir une bonne précision quant à la prédiction de la nature physico-chimique du monomère activé.

La compréhension du mécanisme assurant la spécificité des domaines d'adénylation représente une étape cruciale de la compréhension des NRPS et de leur produit. En effet, la découverte de la correspondance entre les domaines d'adénylation et le monomère activé permet alors de relier une synthétase NRPS à son produit, voire de prédire le produit de synthétases. Le code de Stachelhaus et la méthode à base de SVM de Rausch *et al.* sont, à ce jour, toujours utilisés dans les outils de prédiction et d'analyse des produits NRPS et sont présentés dans ce manuscrit (paragraphe 1.1.6).

#### 1.1.4.2 Domaine de condensation

Les domaines de condensation (domaine C) sont principalement responsables de la formation d'une liaison peptidique entre deux monomères activés par les domaines A de 2 modules successifs. Néanmoins, l'étude de ces domaines a démontré qu'il était possible de distinguer différents types de domaines de condensation en fonction de la nature du substrat impliqué dans la réaction (Rausch *et al.*, 2007, Caradec *et al.*, 2014). Ainsi, les domaines  $^L C_L$  sont responsables de la condensation de deux acides aminés de la série L. Les domaines  $^D C_L$ , présents au sein de module suivant un domaine d'épimérisation (domaine E), sont responsables de la liaison entre l'acide aminé de la série D du module précédent et l'acide aminé de la série L du domaine auquel ce domaine  $^D C_L$  appartient. Les domaines E ont également été identifiés comme des sous-types de domaine C, ainsi que les domaines Dual C/E, responsables de la modification de l'isomérisation d'un monomère et de la condensation de ce monomère avec le suivant dans les synthétases de lipopeptides chez *Pseudomonas* (Balibar *et al.*, 2005). L'existence de ces domaines a été mise en évidence par l'identification de NRPS produisant des peptides contenant des acides aminés de la série D sans qu'aucun domaine E ne soit détecté au sein de la synthétase. Au début de la chaîne d'assemblage se trouve un module d'initiation généralement composé d'un domaine A pour l'activation du premier monomère du peptide, et d'un domaine T. Dans certains cas, un domaine C starter peut également être présent, ce type de domaine C étant impliqué dans la liaison entre le premier monomère activé avec un acide gras dans le cas de la synthèse d'un lipopeptide notamment.

Enfin, des domaines de condensation situés à l'extrémité C-terminale de synthétase d'origine fongiques ont été détectés. Leur activité a été étudiée récemment *in vitro*, et ces domaines ont été démontrés comme étant responsables de la macrocyclisation des peptides non ribosomiques produits par ces synthétases (Gao *et al.*, 2012) (Figure 5).

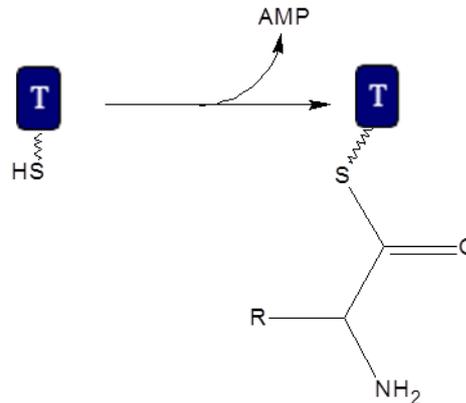


**Figure 5 : Variété des domaines de condensation présents dans les NRPS. La division du domaine C en deux sous-parties est présentée. La partie « Down-seq », composée des 150 derniers acides aminés du domaine détermine son activité enzymatique (Caradec *et al.*, 2014).**

L'identification de sous-types de domaines de condensation est importante dans le cadre de l'étude des NRPS et de leur produit. En effet, la détection d'un domaine C d'un certain sous-type peut fournir une information importante sur une modification du monomère activé par le module portant ce domaine, ou sur la structure globale du peptide. Ainsi, la présence d'un domaine du type  ${}^D C_L$  indique l'activation d'un acide aminé de la forme D- par le module précédant celui portant ce type de domaine C. La présence d'un domaine Ct à l'extrémité d'une synthétase identifiée chez une moisissure signifie que le peptide produit par la synthétase sera probablement cyclisé entre le premier et le dernier acide aminé.

#### 1.1.4.3 Domaine de thiolation

Les domaines de thiolation (domaine T, ou PCP pour Peptidyl Carrier Protein) assurent au sein de chaque module la liaison entre la synthétase et le peptide en formation. Cette liaison est effectuée grâce à une réaction de thiolation entre une phosphopantéthéine portée par le domaine T et l'acide aminé activé par le domaine d'adénylation (Figure 6).



**Figure 6 : Schéma de l'activité enzymatique du domaine de thiolation**

Pour être actif, un domaine de thiolation doit porter un co-facteur acyl-CoA doit être lié au domaine de thiolation, permettant de passer de la forme apoenzyme à la forme holoenzyme inactive.

#### 1.1.4.5 Domaine de thioestérase

A l'extrémité C-terminale de la synthétase se trouve un module de terminaison composé d'un domaine C, un domaine A, un domaine T et d'un domaine de thioestérase (domaine Te). Ce domaine est responsable du relargage du peptide par clivage enzymatique de la liaison thioester reliant le peptide en formation au dernier domaine T de la synthétase. Le domaine Te peut également être responsable de la macrocyclisation du peptide, si celui-ci est cyclique ou partiellement cyclique (Figure 7).

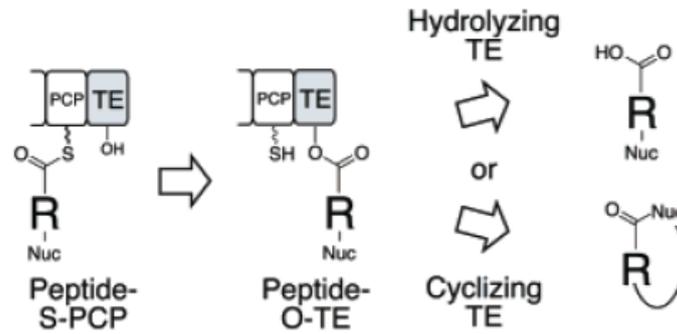


Figure 7 : Schéma du mécanisme d'action de la thioestérase. Après le transfert du peptide en formation du domaine T au domaine Te, la thioestérase catalyse le relargage du peptide ou sa cyclisation (Kohli *et al.*, 2001).

Un second domaine Te est parfois retrouvé dans certaines NRPSs. Le rôle de cette deuxième thioestérase est longtemps resté obscur, mais l'étude *in vivo* de son activité a démontré que son rôle serait la régénération des domaines T chargés avec un mauvais cofacteur, n'aboutissant pas à la formation de l'holoenzyme, rendant alors la synthétase non fonctionnelle (Schwarzer *et al.*, 2002. Roongsawang *et al.*, 2007) (Figure 8).

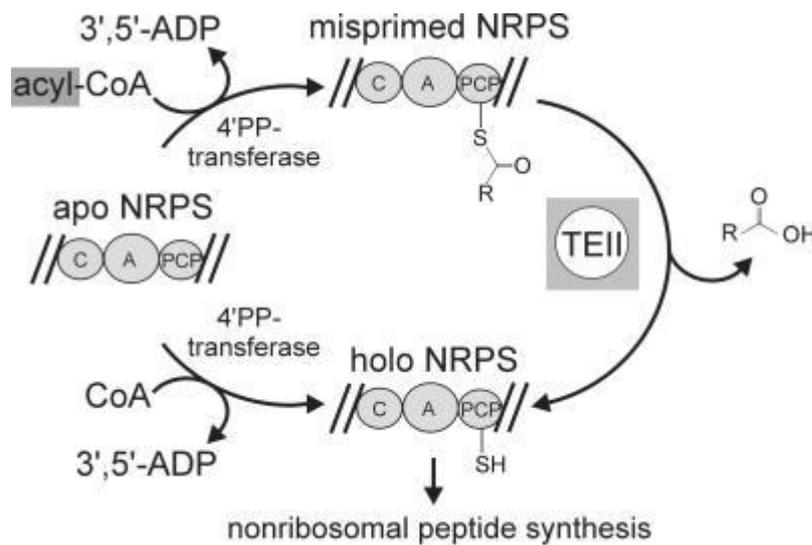


Figure 8 : Mécanisme de réparation des domaines de thiolation par la TeII, permettant le remplacement du mauvais co-facteur chargé par le domaine T (acyl-CoA) par un CoA, permettant d'obtenir un domaine T fonctionnel.

#### 1.1.4.6 Domaines de modification

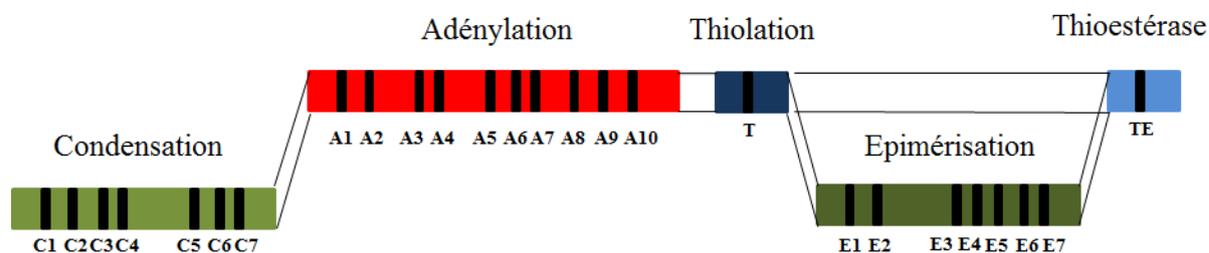
D'autres domaines enzymatiques, responsables de la modification des monomères du peptide au cours de sa synthèse, peuvent également être retrouvés au sein des modules NRPS.

Parmi ceux-là, le domaine le plus fréquemment rencontré est le domaine d'épimérisation (domaine E), responsable du changement d'isomérisation L en une forme D du monomère activé par le module contenant ce type de domaine. Le domaine de méthylation (domaine M) est responsable de la méthylation d'un atome d'azote.

De la même façon que pour les domaines essentiels (notamment les domaines A et C), l'identification et la compréhension du mécanisme de fonctionnement des domaines secondaires sont particulièrement importants pour définir ou prédire le produit d'une NRPS identifiée.

#### 1.1.4.7 Identification des domaines enzymatiques : signatures

Les différents domaines composant les NRPS possèdent des séquences conservées en acides aminés réparties sur la longueur du domaine, appelées signatures (Marahiel *et al.*, 1997). Par exemple, le domaine d'adénylation possède 10 séquences conservées, notées A<sub>1</sub> à A<sub>10</sub>, permettant l'identification des domaines A des NRPS (Figure 9). L'existence de ces différentes signatures peut donc être utilisée pour l'identification des différents domaines enzymatiques NRPS. Si le rôle de ces séquences conservées n'a pas été identifié pour chacune d'entre elles, certaines semblent jouer un rôle crucial dans le fonctionnement des synthétases. Ainsi, la séquence conservée A9 joue un rôle important dans la présentation du monomère activé par le module portant ce domaine au domaine T suivant (Bučević-Popović *et al.*, 2012). De plus, ces signatures possèdent une variabilité entre les organismes procaryotes et eucaryotes (Lee *et al.*, 2005)



**Figure 9 :** Représentation schématique des séquences conservées (signatures) des domaines NRPS les plus communs (Marahiel *et al.*, 1997).

#### 1.1.5 Diversité des peptides issus de la voie non ribosomique

Le mode de fonctionnement des NRPS, de par la possibilité de diverses modifications des acides aminés composant les peptides formés, ou par l'intégration d'acides aminés non protéogéniques, aboutit à une grande diversité de structures de peptides, et une grande diversité d'activités biologiques.

### 1.1.5.1 Diversité de structures des peptides non-ribosomiques

Les produits de la voie non-ribosomiques appartiennent à plusieurs catégories de structure. Les trois types de structure les plus retrouvés sont des structures linéaires (26,6%), cycliques (38,1%) et partiellement cycliques (26,5%). Les peptides branchés (0,4%) et bicycliques (2,2%) ne représentent qu'une faible proportion des peptides non-ribosomiques (données issues de Norine, [bioinfo.lifl.fr/norine](http://bioinfo.lifl.fr/norine)).

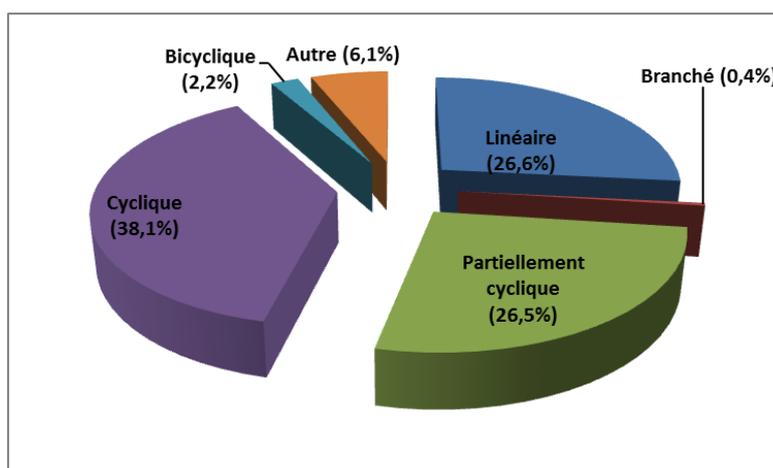
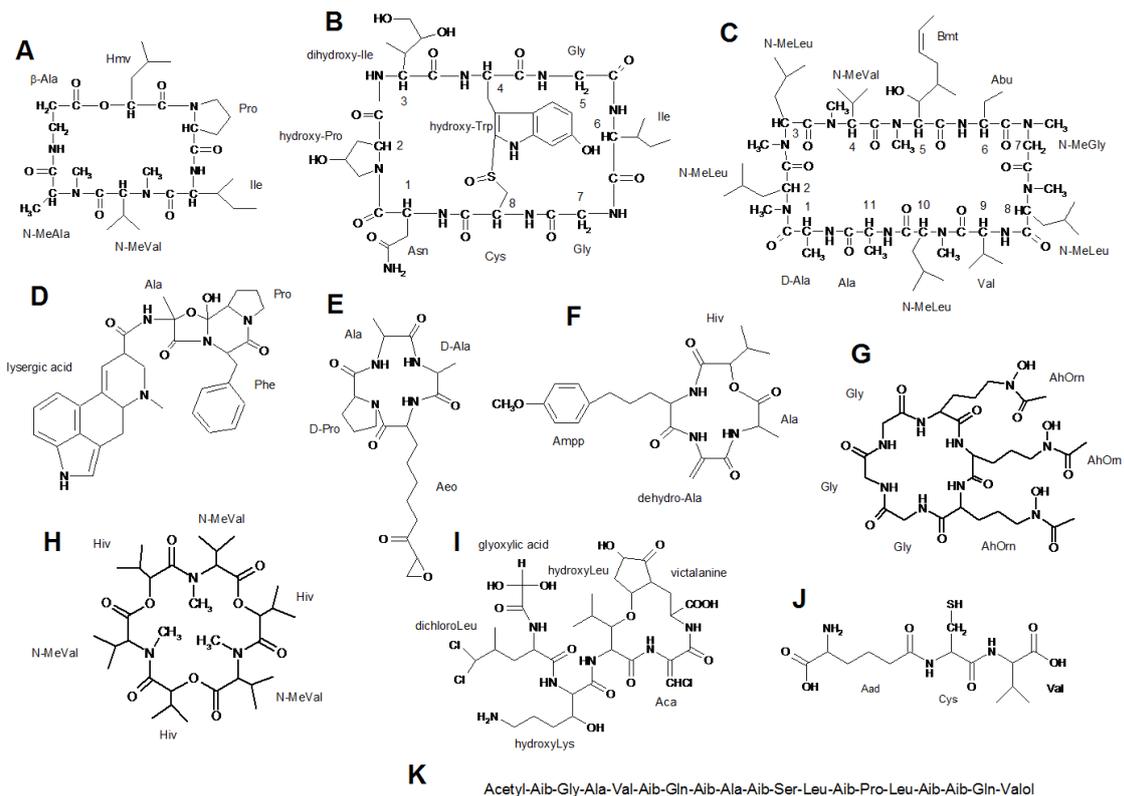


Figure 10 : Structure des peptides non-ribosomiques

En plus des différents types de structures, les NRP présentent également une grande variété de nature. Ainsi, on retrouve parmi ces molécules des peptides, composés simplement d'une chaîne d'acides aminés, des lipopeptides, portant un acide gras, des glycopeptides, des chromopeptides ou des peptaibols (Figure 11).



**Figure 11 : Représentation de la diversité structurelle des molécules issues de la voie non-ribosomique. (A : Destruxine ; B : Amanitine ; C : Cyclosporine ; D : Ergotamine ; E : HC-toxine ; F : AM-Toxine ; G : Ferrichrome ; H : Enniatine ; I : Victorine ; J : ACV ; K : Peptaibol)**

Le mécanisme NRPS permet donc d'aboutir à la production d'une grande variété de structures et de nature de molécules, pouvant être composées d'acides aminés habituellement peu retrouvés dans les molécules directement issues de la traduction d'ARNm avec ou non des cyclisations complètes ou partielles, ou des condensations de la chaîne peptidique avec un acide gras ou un chromophore.

Cette variété de structure entraîne alors une grande variété des activités des molécules NRPS, et illustre l'énorme potentiel représenté par la découverte, la compréhension et la maîtrise de la voie de synthèse non-ribosomique.

### 1.1.5.2 Gamme des activités des peptides non-ribosomiques

Les activités antibiotiques et toxiques sont les plus représentées parmi les peptides référencés, représentant respectivement 44,2% et 18,1% des activités. D'autres activités sont également retrouvées : surfactantes (9,6%), antitumorales (7,0%) et de sidérophore (6,2%) (Figure 12).

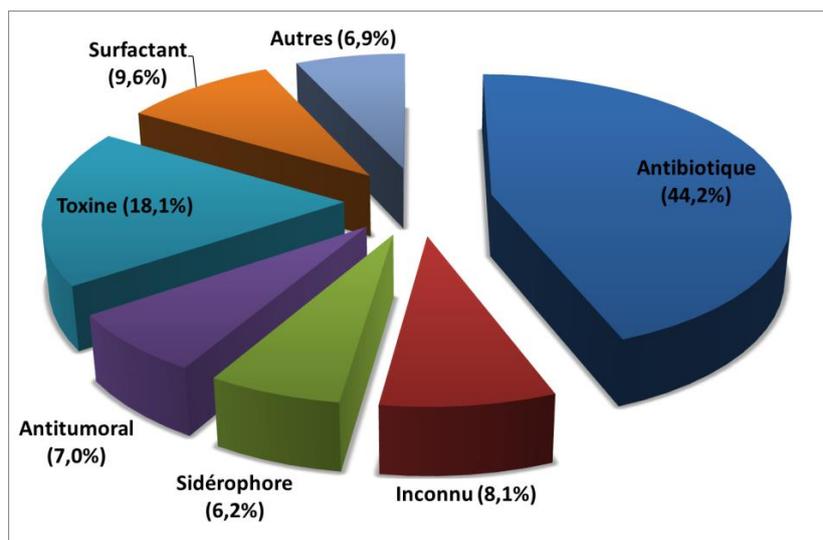


Figure 12 : Activités des peptides non ribosomiques référencés dans NORINE

Les différentes activités des peptides non ribosomiques permettent d'envisager de nombreuses voies de valorisation possibles. Ainsi, plusieurs peptides issus de la voie de synthèse NRPS sont d'ores et déjà utilisés dans l'industrie pharmaceutique. Parmi les plus connus, la pénicilline, découverte par Alexander Fleming en 1918, a pour précurseur l'ACV (delta-L-(alpha-amino adipoyl)-L-cysteinyl-D-valine), un tripeptide produit notamment par le genre *Penicillium*, dont la synthétase est codée par le gène *pcbAB*, ou la gramicidine S, produite par *B. brevis*.

Les surfactants produits peuvent également être valorisés dans plusieurs domaines industriels, notamment les domaines de la cosmétique et de l'agroalimentaire, pour leurs propriétés tensioactives et émulsifiantes. Ceux-ci possèdent l'avantage de représenter une alternative d'origine naturelle aux molécules utilisées actuellement pour ces applications, souvent issues de la synthèse chimique. Parmi les molécules surfactantes d'origine NRPS, la

surfactine, produite par *Bacillus subtilis*, possède des propriétés tensioactives particulièrement intéressantes.

#### *1.1.6 Outils bioinformatiques pour l'étude des NRPS*

De par la particularité du mode de synthèse non ribosomique, les outils classiques d'études de séquences génomiques ne permettent pas l'analyse complète des NRPS, et ne permettent notamment pas la prédiction des peptides produits par ces synthétases. Le développement d'outils spécifiques a donc été requis pour pouvoir analyser les séquences codant pour des NRPS au sein de génomes microbiens.

En raison de l'augmentation rapide du nombre de génomes séquencés à travers le monde, les outils de criblage bioinformatique doivent permettre une détection rapide de gènes potentiellement impliqués dans la production de métabolites secondaires. Une prédiction précise de la structure des molécules produites est également nécessaire. Les connaissances accumulées depuis la découverte du mécanisme de synthèse non-ribosomique concernant le rôle des différents domaines et la spécificité des domaines d'adénylation ont permis la mise au point d'outils dédiés à l'identification et la caractérisation des NRPS à partir de séquences génomiques ou protéiques et à la prédiction de leur produits.

Les outils globaux d'analyse de génome et de domaines enzymatiques sont présentés ici, ainsi que les outils spécifiquement dédiés à l'analyse des NRPS et de leurs produits.

#### **NRPS-PKS** (<http://www.nii.res.in/nrps-pks.html>)

Développé par une équipe de l'Institut National d'Immunologie de New Dehli, NRPS-PKS permet, à partir d'une séquence protéique, de détecter et d'identifier la présence de domaines NRPS (A, T, C, E et Te). Le logiciel permet l'identification des domaines NRPS par le biais d'une comparaison de la séquence à analyser avec les séquences d'une base de données interne (NRPSDB). Le portail fournit le résultat sous la forme d'une représentation schématique de l'organisation en domaines enzymatiques identifiés et permet de récupérer les séquences protéiques les composant (Ansari *et al.*, 2004).

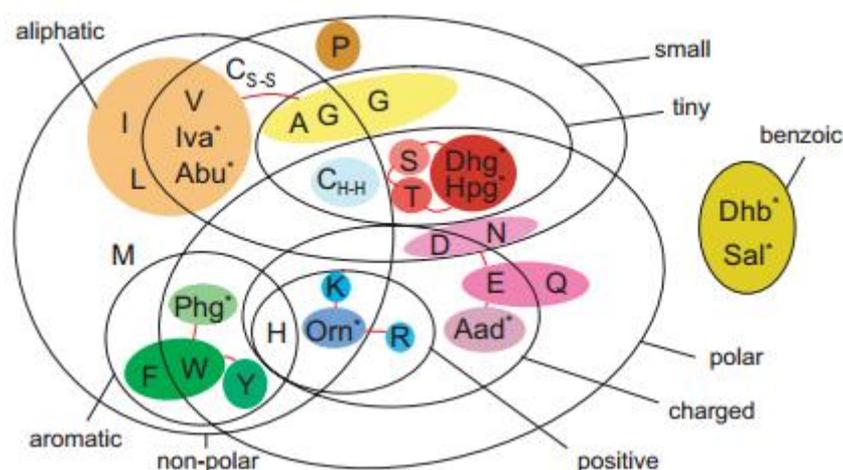
#### **PKS/NRPS Analysis** (<http://nrps.igs.umaryland.edu/nrps/>)

Comme NRPS-PKS, cet outil permet à partir d'une séquence protéique d'identifier les domaines enzymatiques composant les synthétases de peptides non ribosomiques et les

synthèses de polykétides, et les représente sous forme schématique. La récupération des séquences composant les domaines est également possible.

**NRPS Predictor 2** (<http://nrps.informatik.uni-tuebingen.de/>)

Le portail NRPS Predictor 2 est dédié à la prédiction des monomères activés par les domaines A de NRPS. Le logiciel réalise des prédictions sur 4 niveaux de précision différents. Le premier niveau ne concerne qu'une prédiction des propriétés physico-chimiques du substrat du domaine étudié (Figure 13). Les deuxième et troisième niveaux de prédictions permettent de déterminer un groupe d'acides aminés auquel appartient le monomère activé. Dans le premier cas (modèle « Large Cluster »), le modèle permet de classer le monomère dans un groupe de 1 à 7 acides aminés, et dans le deuxième cas (modèle « Small Cluster »), dans un groupe de 1 à 5 acides aminés. Enfin, le dernier niveau de prédiction, « Single AA », effectue la prédiction directe du monomère activé. Naturellement, plus le niveau de détail de la prédiction est important, plus la prédiction possède une précision faible, notamment dans le cadre de la prédiction d'acides aminés peu communs dans les peptides non-ribosomiques (Röttig *et al.*, 2011).



**Figure 13 : Regroupement des acides aminés selon leurs propriétés physico-chimiques utilisées pour les prédictions réalisées par NRPS Predictor 2 avec les modèles Large Cluster et Small Cluster**

La prédiction réalisée dans le cadre du premier modèle, en se basant sur les propriétés physico-chimiques du monomère activé par le domaine d'adénylation concerné, trouve un intérêt lorsque la prédiction réalisée avec le modèle « Single AA » ou « Nearest Neighbor » ne permet pas d'obtenir une prédiction avec la précision suffisante.

Pour effectuer ces prédictions, l'apprentissage a été réalisé en utilisant un set de plus de 4000 domaines d'adénylation dont les résidus formant les sites actifs correspondant aux 34 acides aminés compris dans une poche de 8Å autour du substrat du domaine.

### AntiSMASH

AntiSMASH (Antibiotics & Secondary Metabolites Analysis SHell) est un portail bioinformatique dédié à la détection et l'analyse de clusters de gènes impliqués dans la production de métabolites secondaires (Medema et *al.*, 2011). Ce portail permet l'analyse de séquences d'ADN de génomes entiers ou de drafts, sous la forme d'un fichier GenBank, EMBL ou FASTA. Les résultats sont présentés sous la forme de clusters numérotés auxquels un type est attribué. A ce jour, 24 types de clusters de gènes peuvent être identifiés en fonction de leurs produits, dont les NRPS, les PKS, les lantibiotiques, les sidérophores ou les homosérines lactones (HSL). Pour chaque cluster, les différents gènes sont représentés et un rôle leur est attribué, à savoir les gènes directement impliqués dans la biosynthèse du métabolite, les gènes impliqués dans le transport du produit final, et les gènes de régulation du cluster. Le contexte génomique du cluster détecté est également présenté, avec la représentation des autres gènes situés à proximité du cluster, mais n'étant pas directement impliqués dans la synthèse d'un métabolite secondaire. Dans le cas des NRPS, l'organisation en domaines enzymatiques de la synthétase est représentée pour chaque gène de biosynthèse. Une prédiction des monomères activés par chaque domaine A est également réalisée, notamment au moyen de NRPS Predictor 2 et du code de Stachelhaus. Un consensus issu des 3 prédictions est alors déterminé, et une structure globale du peptide prédit est présentée. Enfin, une recherche de clusters apparentés au cluster détecté est réalisée et présentée, et permet d'identifier des clusters présents chez d'autres organismes dont l'organisation des gènes de production est proche.

## Comparaison des logiciels d'analyse NRPS

Il existe donc plusieurs logiciels dédiés à l'étude des NRPS. Chaque logiciel possède ses propres particularités, et ses propres performances.

**Tableau 1 : Comparaison des logiciels destinés à l'étude de NRPS.**

Logiciel	Format	Résultats
NRPS-PKS	Protéine	Modules et domaines NRPS
PKS-NRPS Analysis	Protéine	Domaines NRPS
NRPS Predictor 2	Protéine	Acides aminés activés par les domaines A
AnsiSMASH	ADN, Protéine	Clusters NRPS identifiés, organisation en domaines, identification de gènes de régulation et de transport, alignement avec des clusters de gènes NRPS proches. Prédiction des monomères activés.

Chaque logiciel est donc utilisé en fonction du résultat recherché. Une analyse totale du potentiel de production NRPS d'une souche bactérienne à partir de son génome complet n'est par exemple directement réalisable qu'avec antiSMASH, et ce en quelques heures. En revanche, l'identification simple des domaines enzymatiques composant une synthétase identifiée peut être effectuée très rapidement au moyen des logiciels NRPS-PKS et PKS-NRPS Analysis, qui fournissent un résultat en quelques secondes. Enfin, la prédiction d'un peptide produit par une synthétase identifiée peut être réalisée au moyen du logiciel NRPS Predictor 2, fournissant en quelques minutes les prédictions réalisées au moyen des différents modèles de prédictions utilisés.

Il est important de garder à l'esprit que ces outils sont des outils prédictifs, et ne sauraient être utilisés pour obtenir une réponse définitive quant au rôle exact de chaque domaine enzymatique identifié, ou la prédiction des acides aminés activés par les domaines d'adénylation. La prise en compte du score associé à chaque prédiction, permettant d'évaluer la qualité de chacune d'entre elles, permet d'utiliser ces outils avec la prudence nécessaire, sans oublier les limites inhérentes à chacun d'entre eux.

**NORINE** (<http://bioinfo.lifl.fr/norine/>)

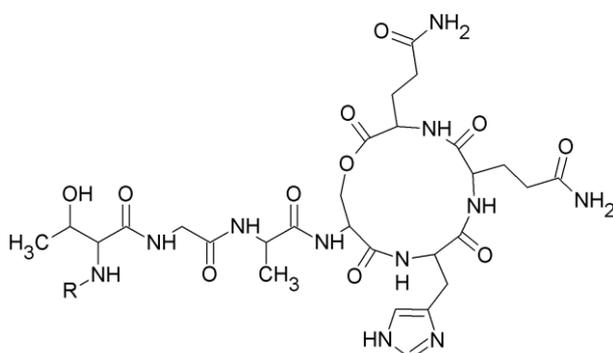
NORINE est à ce jour la seule base de données dédiée aux peptides d'origine non ribosomique. Développée dans le cadre d'une collaboration entre le laboratoire ProBioGEM et le LIFL (Laboratoire d'Informatique Fondamentale de Lille) de Lille1, et comprenant actuellement environ 1150 molécules, cette base de données donne pour chaque peptide les informations disponibles sur celui-ci : les données structurales, les organismes producteurs identifiés, les activités biologiques de la molécule et les informations sur la synthétase responsable de la production du peptide lorsque cette information est disponible. NORINE permet de rechercher un peptide de façon générale par son nom ou sa composition globale et de chercher des peptides par structure complète ou partielle (Caboche *et al.*, 2008).

## 1.2 La kurstakine

Découverte en 2000, la kurstakine fait partie des lipopeptides produits par le genre *Bacillus* comme la surfactine ou la fengycine. La kurstakine est à ce jour la famille de lipopeptides de *Bacillus* la plus récemment découverte. L'objectif de cette partie est de réaliser un état de l'art sur les connaissances acquises sur le lipopeptide en termes de structure, de mécanisme de synthèse et d'activité biologique.

### 1.2.1 Structure et découverte

La kurstakine est un lipopeptide identifié pour la première fois chez *Bacillus thuringiensis kurstaki* HD-1 (Hathout *et al.*, 2000). C'est un lipopeptide, composé des acides aminés Thr-Gly-Ala-Ser-His-Gln-Gln, et d'une chaîne acide gras de 11 à 14 carbones (Figure 14). L'acide gras porté peut être  $\beta$ -hydroxylé. La kurstakine peut être cyclisée *via* une liaison lactone entre la sérine en position 4 et la glutamine en position 7.



Acide gras (R)	Linéaire	Cyclique
iC11	ND	877,465
C12	ND	891,481
C12 [-OH(3)]	925,486	907,476
iC12	ND	891,481
iC13	ND	905,496
iC13 [-OH(3)]	939,502	921,491
C14 [-OH(3)]	953,51	935,507

Figure 14 : Structure de la kurstakine et masse moléculaire en fonction de la taille de l'acide gras (R). ND : Non détecté.

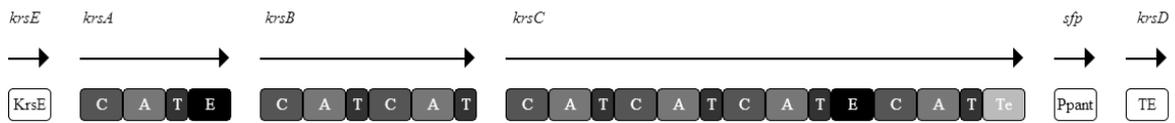
La première identification de la kurstakine a été effectuée à partir de cellules de *B. thuringiensis* HD-1 cultivées sur un milieu de sporulation, le milieu AK agar. Les formes isoC11 (m/z 879), C12 et isoC12 (m/z 893) et isoC13 (m/z 907) ont alors été séparées par HPLC et analysées par spectrométrie de masse (Hathout *et al.*, 2000). Des peptides de masse m/z 892, 906 et 920 ont été détectés quelques années plus tard chez une souche de *B. subtilis* ATCC6051 cultivée sur gélose Trypticase Soja, correspondant aux formes iC12, iC13 et iC13 [-OH(3)] de la kurstakine (Madonna *et al.*, 2003). Des kurstakines de masse m/z 925, 939 et 953 ont également été identifiées en cultivant une souche de *Bacillus sp.* NK2018 pendant 9 jours dans du milieu minimum liquide M9. La différence de masse de 18 Da avec les formes

$\beta$ -hydroxylées de kurstakines C12, iC13 et C14 suggère une différence au niveau de la structure globale du lipopeptide qui ne serait donc pas cyclique comme les formes de kurstakines détectées auparavant (Bumpus *et al.*, 2009). La détection de kurstakines dans le surnageant de la culture en milieu M9 n'a été réalisable qu'en spectrométrie de masse et leur dosage n'a pas été réalisé, suggérant qu'elles étaient présente en très faibles quantités.

### 1.2.2 Mécanisme de synthèse

Deux approches ont été utilisées pour la détection des gènes impliqués dans la production de la kurstakine. La première approche consistait en la détection par spectrométrie de masse permettant de détecter les synthétases et les ions uniques dérivés du cofacteur ppant (Bumpus *et al.*, 2009). Une deuxième approche consistait en une détection par utilisation d'amorces PCR dégénérées, définies à partir de gènes NRPS connus (Tapi *et al.*, 2010). Ces approches ont permis d'identifier la kurstakine synthétase et les gènes la codant.

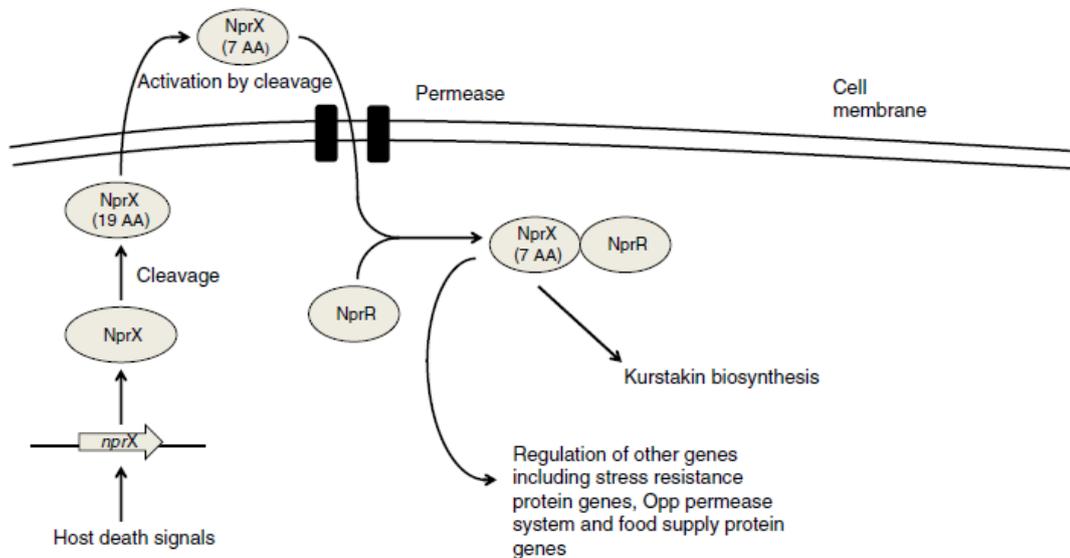
La kurstakine est ainsi issue de la voie NRPS. Les gènes *krsE*, *krsA*, *krsB*, *krsC*, *sfp* et *krsD* codant la kurstakine synthétase ont été identifiés pour la première fois en 2009 (Figure 15). Les gènes *krsA*, *krsB* et *krsC* codent la kurstakine synthétase. *krsA* code pour un module (C-A-T-E), responsable de l'activation et l'incorporation de la thréonine en position 1, *krsB* pour 2 modules (C-A-T-C-A-T) pour l'activation et l'incorporation de la glycine et de l'alanine en positions 2 et 3 respectivement, et *krsC* pour 4 modules (C-A-T-C-A-T-C-A-T-E-C-A-T-Te) pour l'activation de la sérine, la glycine et de 2 glutamines aux positions 4 à 7. La présence d'un domaine d'épimérisation dans le premier module codé par *krsA* et le troisième module codé par *krsC* indique la modification de l'isomérisation du monomère activé par le domaine A de ces modules, menant à la présence d'une D-Thréonine en position 1 et d'une D-Glutamine en position 6. Néanmoins, cette isomérisation de forme D n'a jamais été démontrée expérimentalement. Le domaine Te du dernier module codé par *krsC* est responsable de la cyclisation du peptide synthétisé et du clivage de la liaison thioester le reliant à la synthétase. Le gène *krsD* est responsable de la production d'une seconde thioestérase de type II dont le rôle précis n'a pas été élucidé à ce jour. Le gène *sfp* est responsable de la production d'une phosphopantéthéinyl transférase. Le gène *krsE* situé en amont des gènes de synthétase est potentiellement impliqué dans la production d'une protéine d'efflux, pouvant être responsable de la sécrétion de la kurstakine dans le milieu.



**Figure 15 : Organisation des gènes et des domaines de la kurstakine synthétase. Les flèches représentent les gènes et les boîtes les domaines enzymatiques (Bumpus *et al.*, 2009).**

### 1.2.3 Régulation

Le mécanisme de régulation de production de la kurstakine synthétase et du lipopeptide associé est à ce jour peu connu. Néanmoins, les gènes de la kurstakine synthétase ont été identifiés comme faisant partie d'un groupe de gènes sous l'influence du régulateur NprR, impliqué dans la régulation de gènes de virulence permettant la survie des cellules de Bt dans des cadavres d'insectes infectés (Dubois *et al.*, 2012) (Figure 16).



**Figure 16 : Mécanisme de régulation de la synthèse de la kurstakine. La mort de l'insecte infecté provoque la transcription du gène *nprX*, dont le produit est activé par deux clivages successifs, le deuxième étant réalisé de façon extracellulaire. Le peptide NprX actif peut, après sa réintégration au sein de la cellule de Bt via une perméase, s'associer au régulateur NprR, le complexe NprR-NprX régulant alors la transcription d'un ensemble de gènes dont font partie les gènes *krsABC* de la kurstakine synthétase (Béchet *et al.*, 2012).**

Le régulateur NprR agit sur la transcription d'un ensemble de 41 gènes impliqués notamment dans la production de protéines de stress, de perméases, de protéases et de transports de nutriments. Parmi les gènes dont la transcription est affectée par le régulateur

NprR se trouvent les gènes *krsABC* de la kurstakine synthétase. Le régulateur NprR est activé à la suite d'une cascade de réactions aboutissant à la transcription du gène *nprX*, et le clivage par activation du peptide produit, permettant alors la formation d'un complexe NprR-NprX, responsable de la modification de la régulation des 41 gènes. La transcription de *nprX* étant activée à la mort de l'hôte insecte infecté par Bt, la kurstakine jouerait ainsi un rôle dans le mécanisme nécrotrophique décrit chez *B. thuringiensis* (Dubois *et al.*, 2011).

#### 1.2.4 La kurstakine, un lipopeptide intracellulaire ?

A ce jour, les kurstakines n'ont donc pu être dosées qu'à partir de cultures réalisées en milieu solide, indiquant que ce lipopeptide est peut-être peu excrété dans le milieu de culture s'il est liquide, ou est lié à la paroi cellulaire des bactéries productrices. Cette particularité marque une différence importante avec les autres lipopeptides produits par le genre *Bacillus*, qui sont généralement retrouvés dans le milieu de culture.

Comme présenté dans le paragraphe 1.2.2, le gène *krsE* présent en amont des gènes de la kurstakine synthétase *krsABC* code potentiellement pour une protéine d'efflux. Néanmoins, le rôle exact de cette protéine n'a pas été démontré expérimentalement. Si la protéine KrsE a bien pour rôle l'excrétion de la kurstakine dans le milieu extracellulaire, la transcription du gène correspondant pourrait être un facteur limitant dans la production de kurstakine en milieu liquide. Il est également important de noter qu'à ce jour, il n'existe pas de données précises sur les promoteurs associés aux gènes de la kurstakine synthétase. Ainsi, le gène *krsE* pourrait être sous la dépendance d'un promoteur différent de ceux des gènes *krsABC*, et donc aurait un mécanisme de régulation différent.

#### 1.2.5 La kurstakine, biomarqueur de *Bacillus cereus*

La production de la kurstakine au sein du genre *Bacillus* a été étudiée par spectrométrie de masse MALDI-ToF par l'analyse de 54 isolats de ce genre à la recherche des différentes familles de lipopeptides connues. Parmi les 54 souches analysées, la production du lipopeptide a été détectée chez 20 d'entre elles (Price *et al.*, 2007).

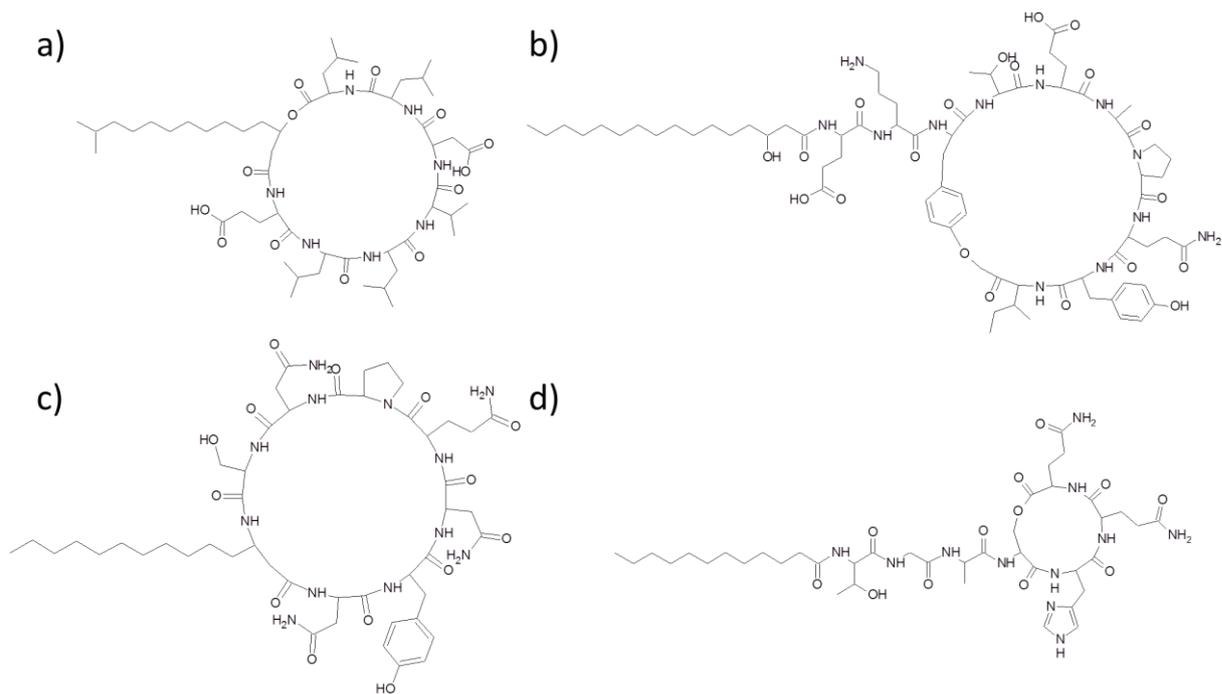
Le potentiel de production des souches du genre *Bacillus*, a quant à lui, été analysé par l'analyse bioinformatique du génome (partiel ou complet) de 44 souches de *Bacillus*. Les gènes de la kurstakine synthétase ont été retrouvés chez 32 souches, appartenant toutes au

groupe *Bacillus cereus*. Les gènes n'ont toutefois pas été détectés chez *B. anthracis* et *B. thymototoxicus*, malgré leur appartenance au groupe *cereus* (Béchet *et al.*, 2012).

A ce jour, la kurstakine n'a ainsi été retrouvée qu'au sein du genre *Bacillus*, et du groupe *cereus* en particulier. Cette spécificité pourrait conférer à ce lipopeptide le statut de biomarqueur pour ce groupe. De plus, la détection de production de kurstakine et des gènes de la kurstakine synthétase chez des souches de *Bacillus* n'appartenant pas à l'espèce *thuringiensis* indique que le lipopeptide ne peut pas en revanche servir de biomarqueur pour l'espèce *thuringiensis*, dont la différenciation et la séparation avec l'espèce *cereus* reste difficile, semblant indiquer que ces deux espèces ne devraient être considérées que comme une seule (Helgason *et al.*, 2000 ; Chen et Tsen, 2002).

#### *1.2.6 Les différentes familles de lipopeptide produits par Bacillus*

Le genre *Bacillus* est connu pour la production d'une grande variété de lipopeptides d'origine non-ribosomique regroupés en différentes familles : les surfactines, les fengycines, les iturines et les kurstakines (Jacques, 2011). Ces différentes molécules sont toutes composées d'une partie peptidique et d'une partie lipidique (Figure 17). Ces lipopeptides sont soit cycliques, le cycle étant alors formé entre le premier et le dernier acide aminé de la partie peptidique, comme pour la surfactine et l'iturine, soit partiellement cycliques, le cycle étant alors formé par deux acides aminés de la chaîne peptidique : le dernier acide aminé de la chaîne et un autre acide aminé, comme c'est le cas chez la fengycine et la kurstakine.



**Figure 17 : Structures des 4 grandes familles de lipopeptides du genre *Bacillus* : La surfactine (a), la fengycine (b), l'iturine (c) et la kurstakine (d) (Jacques, 2001).**

Ces différentes familles regroupent les lipopeptides en fonction du nombre d'acides aminés qui les composent et de la nature et l'isomérisation de ces derniers. Ainsi, la famille des surfactines regroupe la surfactine, la pumilacidine, la lichenysine, l'espérine et la bamylocine A. La famille des iturines regroupe l'iturine, la bacillomycine et la mycosubtiline. Enfin, la famille des fengycines regroupe la fengycine et la plipastatine (Tableau 2).

**Tableau 2 : Récapitulatif des structures globales des familles de lipopeptides produits par le genre *Bacillus***

<b>Nom</b>	<b>Structure peptidique</b>	<b>Acide gras</b>
<b>Surfactine</b>	Glu-(Val/Leu/Ile)-Leu-(Aal/Val/Leu/Ile)-Asn-Leu-(Val/Leu/Ile)	C <sub>14</sub> , C <sub>15</sub>
<b>Fengycine</b>	Glu-Orn-Tyr-Thr-Glu-(Ala/Val)-Pro-Gln-Tyr-Ile	C <sub>15</sub> , C <sub>16</sub> , C <sub>17</sub>
<b>Iturine</b>	Asn-Tyr-Asn-(Pro/Gln/Ser)-(Glu/Pro)-(Ser/Asn)	C <sub>14</sub> , C <sub>15</sub> , C <sub>16</sub> , C <sub>17</sub>
<b>Kurstakine</b>	Thr-Gy-Ala-Ser-His-Gln-Gln	C <sub>11</sub> , C <sub>12</sub> , C <sub>13</sub> , C <sub>14</sub>

### 1.2.7 Activités biologiques des lipopeptides de *Bacillus*

Ces molécules possèdent pour la plupart des propriétés valorisables. Ainsi, la surfactine présente un intérêt très important par ses fortes propriétés tensioactives et sa faible Concentration Micellaire Critique (CMC) (Peypoux *et al.*, 1999). Ces caractéristiques, couplées à son origine naturelle et son caractère biodégradable, en font un bon candidat pour le remplacement de surfactants d'origine synthétique utilisés dans les domaines de la cosmétique et de l'agroalimentaire.

La fengycine possède quant à elle des propriétés antifongiques intéressantes contre les champignons filamenteux (Vanittanakom *et al.*, 1986). Une activité immunomodulatrice a également été mise en évidence, notamment pour la plipastatine.

L'iturine présente des propriétés surfactantes moins importantes que la surfactine, mais présente une forte activité antifongique contre une variété de souches de moisissures et de levures. Elle ne possède en revanche pas de propriétés antibactériennes ou antivirales (Mizumoto *et al.*, 2007 ; Romero *et al.*, 2007).

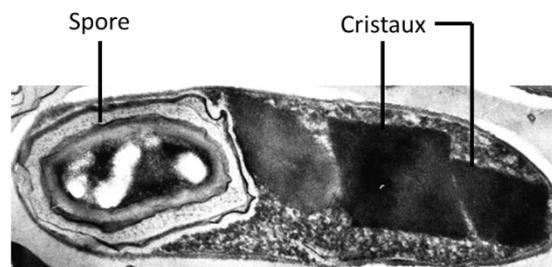
De plus, l'existence d'une synergie de l'activité des lipopeptides a également été mise en évidence, notamment par l'utilisation conjointe de la surfactine et de l'iturine, présentant un effet positif sur le caractère émulsifiant des molécules (Razafindralambo *et al.*, 1997), ainsi que sur le caractère antifongique de l'iturine (Maget-Dana *et al.*, 1992).

A ce jour, peu d'activités biologiques ont été attribuées à la kurstakine. Une inhibition de croissance de la moisissure *Stachybotrys chartarum* par un extrait pur de kurstakine a été observée, mettant en évidence le caractère antifongique de la molécule (Hathout *et al.*, 2000). Plusieurs souches de *B. thuringiensis* ont présenté une activité antifongique contre plusieurs souches de moisissures, mais une corrélation entre cette activité et la production de kurstakine n'a pas pu être clairement établie (Abderrahmani *et al.*, 2011). La diversité des lipopeptides produits par le groupe *Bacillus* et les nombreuses activités biologiques observées soulignent néanmoins le fort potentiel de la kurstakine pour une valorisation industrielle. De plus, le caractère cationique provenant de la présence d'un résidu histidine au sein de la kurstakine pourrait présenter un intérêt important pour l'activité de la molécule, ce caractère ionique étant généralement important pour l'activité antibiotique des peptides ainsi chargés (Jiang *et al.*, 2008).

### 1.2.8 *Bacillus thuringiensis*

#### Généralités

*Bacillus thuringiensis* est un bacille à Gram positif, aérobic et sporulant, appartenant au groupe *Bacillus cereus*, regroupant les espèces *cereus*, *thuringiensis*, *anthracis*, *mycoides*, *weihenstephanensis*, *cytotoxicus*, *toyonensis*, *manliponensis* et *gaemokensis*. Les cellules végétatives de *B. thuringiensis* (communément abrégé Bt) mesurent environ 1µm de largeur et 5µm de longueur.



**Figure 18** Cellule de *Bacillus thuringiensis* en cours de sporulation observée en microscopie électronique. Les cristaux de toxines sont visibles (Ibrahim *et al.*, 2010).

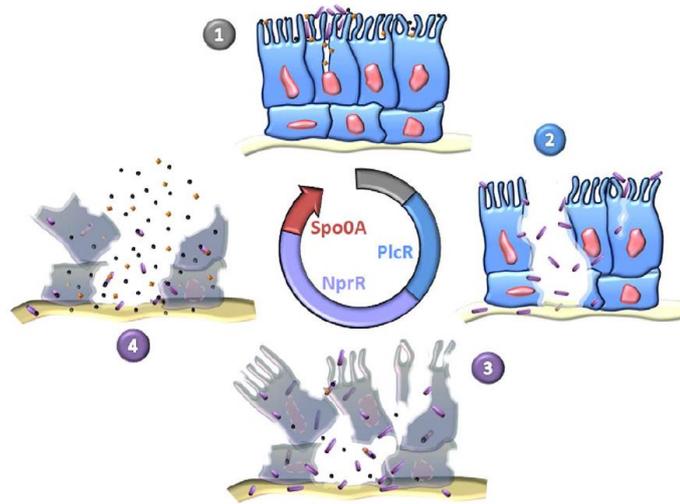
Le premier isolat de *B. thuringiensis* a été réalisé au début du XX<sup>ème</sup> siècle au Japon à partir de vers à soie infectés. Un nouvel isolat sera réalisé par le scientifique Ernst Berliner dans l'état allemand de Thuringe, qui donnera son nom à l'espèce isolée, cette fois à partir de larves de mite de farine (Ibrahim *et al.*, 2010).

#### **Mode de vie de *Bacillus thuringiensis* : Pathogénicité et nécrotrophisme**

*B. thuringiensis* possède un cycle de vie divisé en deux phases distinctes, la première concernant le mode de vie sous forme de cellule végétative, et la deuxième la sporulation de la bactérie. *B. thuringiensis* possède un caractère entomopathogène, particulièrement actif contre les lépidoptères, diptères et coléoptères. Cette pathogénicité est liée à la production d'un cristal parasporal de nature protéique toxique lors de la formation de spores. Le mode d'action de cette toxine repose sur la formation de pores dans la paroi intestinale des insectes après digestion du cristal en endotoxines. Selon les souches, les protéines formant les cristaux produits par *B. thuringiensis* sont codées par une centaine de différents gènes *Cry* et *Cyt*, possédant diverses structures et cibles (Du et Nickerson, 1996 ; Schnepf *et al.*, 1998).

Le mode de vie de *B. thuringiensis* a été décrit comme nécrotrophique en 2012 par Dubois *et al.*, en raison du mode d'action de la bactérie pour l'infection d'insectes cibles, et de la

modification de son profil d'expression génétique une fois l'insecte infecté mort, permettant alors aux cellules de Bt de survivre dans le cadavre de celui-ci, puis de se répandre dans l'environnement jusqu'à la contamination d'un nouvel hôte. Le passage du mode de vie pathogène au nécrotrophisme chez *B. thuringiensis* s'effectue par d'importantes modifications dans le métabolisme des cellules, sous l'influence de différents régulateurs principaux contrôlant les différents stades de l'infection des insectes hôtes (Figure 19).



**Figure 19 : Cycle de pathogénicité de *B. thuringiensis* : Le cycle infectieux est contrôlé par l'expression séquentielle de régulateurs principaux lors des différentes phases de l'infection. 1 : Les spores de Bt et les toxines atteignent l'épithélium intestinal de l'hôte. 2 : Les bactéries produisent les facteurs de virulence régulés par PlcR, attaquant la paroi intestinale. Les bactéries peuvent alors pénétrer les cellules et infecter le sang de l'hôte, provoquant à terme la mort de celui-ci. 3 : Les cellules de Bt passent à un mode de vie nécrotrophique caractérisé par l'expression de gènes régulés par NprR, leur permettant de survivre dans l'hôte mort, puis de sporuler. 4 : Les spores et les toxines se répandent dans l'environnement sous l'effet de facteurs environnementaux. (Dubois *et al.*, 2012)**

Ainsi, le régulateur PlcR régule principalement les gènes liés au mode de vie entomopathogène de *B. thuringiensis*, notamment des gènes impliqués dans la production de toxines et d'enzymes de dégradation (Agaisse *et al.*, 1999). A la mort de l'insecte, le régulateur NprR contrôle l'activité d'un ensemble de 41 gènes contribuant à la survie des cellules de *B. thuringiensis* au sein des cadavres d'insectes, permettant notamment la dégradation enzymatique de protéines, de lipides et de chitine (Dubois *et al.*, 2012). A l'issue de cette phase de nécrotrophisme, le régulateur principal de la sporulation Spo0A permet l'initiation de la production de spores pour la survie de cellules dans des conditions défavorables, notamment l'épuisement de nutriments suite à la mort de l'hôte, et facilite la dissémination des cellules dans l'environnement, jusqu'à l'infection d'un nouvel hôte ou le retour de conditions favorables de croissance.

### ***Bacillus thuringiensis* : un outil de biocontrôle**

A ce jour, le caractère pathogène de *B. thuringiensis* est exploité commercialement de façon importante, en faisant de *B. thuringiensis* l'agent biopesticide le plus utilisé à travers le monde. Bt est classé par la FDA (US Food and Drug Administration) comme GRAS (Generally Recognized As Safe), et les cellules peuvent être utilisées dans l'agriculture biologique pour la protection des cultures particulièrement contre les larves de lépidoptères et de moustiques (Joung et Côté, 2002). A ce jour, les souches utilisées commercialement sont notamment les *B. thuringiensis* sérotype H14, 3A, 3B et H7, ainsi que *B. thuringiensis* sp. Tenebrionis (index phytosanitaire ACTA 2013).

L'utilisation des toxines de *B. thuringiensis* pour la lutte contre d'autres types de parasites tels les nématodes de type Meloidogyne (*Meloidogyne incognita*), agent de la galle végétale a également été envisagée. Des résultats probants ont été obtenus, avec des activités très importantes, présentant un taux de mortalité sur les nématodes testés de 86 à 100% au moyen d'un mélange de spores et de cristaux (Mohammed *et al.*, 2008).

L'utilisation des toxines de *B. thuringiensis* est également connue en raison de l'utilisation des gènes de production de la toxine Cry1Ab par du maïs génétiquement modifié (appelé alors « maïs Bt ») par l'entreprise de biotechnologies agricoles américaine Monsanto.

## 1.3 Les PentaPeptides Cycliques Chlorés (PPCC)

### 1.3.1 Les astines

#### Structure et découverte

Les astines sont une famille de pentapeptides cycliques isolés chez la plante à fleurs *Aster tataricus*, faisant partie des « 50 plantes fondamentales » de la médecine chinoise traditionnelle. Les racines d'*A. tataricus* sont utilisées en décoction pour soigner la toux et l'encombrement des bronches (Schafer, 2011 ; Xinrong, 2003), mais également en comprimé, notamment en mélange avec d'autres plantes médicinales (Figure 20).



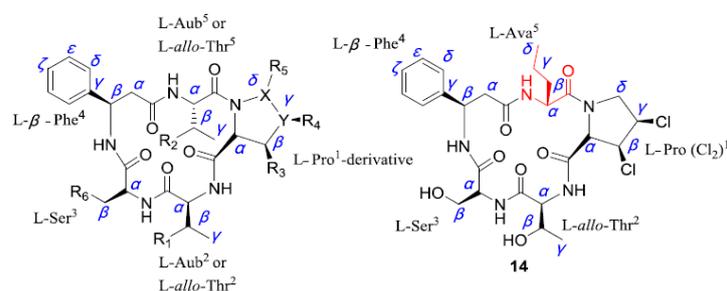
Figure 20 : A. Comprimés à base d'*Aster tataricus* et autres plantes médicinales. B. Racines d'*A. tataricus* utilisées pour les décoctions et extraction d'astines. C. Fleurs d'*A. tataricus*.

Dans le cadre de la problématique de la découverte de nouvelles molécules d'intérêt d'origine naturelle, l'utilisation de ressources anciennes et de méthodes médicales ancestrales représentent une voie de recherche prometteuse pour la valorisation de ces ressources (Gurib-Fakim, 2006). Les racines d'*Aster* sont donc devenues un sujet d'étude pour de nombreuses équipes de recherche pour l'identification de telles molécules.

A ce jour, 16 formes d'astines ont été identifiées (astines A à P). La première molécule identifiée est l'astérine (renommée astine C par la suite), isolée à partir de racines d'*Aster tataricus* (Kosemura *et al.*, 1993). Puis, de nouvelles formes de pentapeptides chlorés ont été identifiées dans un extrait butanolique de racines d'*A. tataricus*, les astines A et B (Morita *et al.*, 1993), et les astines D et E (Morita *et al.*, 1993). Les astines F, G et H sont décrites peu de temps après, toujours extraites à partir de racines de la plante (Itokawa *et al.*, 1994), puis l'astine I (Morita *et al.*, 1994) et enfin l'astine J, présentant la particularité d'être non-cyclique (Morita *et al.*, 1995). La structure de ces molécules a été confirmée lors d'une série de criblages réalisés sur des extraits éthanoliques de racines d'*Aster* par HPLC/DAD/ESI-MS<sup>n</sup>,

permettant également de découvrir de nouveaux composés présents dans ces extraits dont des dérivés linéaires des astines, les asterinines, décrites plus tard dans ce manuscrit (Liu *et al.*, 2012). Enfin, 6 nouvelles formes d'astines (K à P) ont été identifiées à partir d'extraits méthanoliques.

Toutes ces molécules possèdent plusieurs particularités structurales : elles peuvent posséder dans leur chaîne peptidique jusqu'à 4 acides aminés non protéogéniques ( $\beta$ -phénylalanine, *allo*-thréonine (aThr), acide aminobutyrique (Abu) et une proline halogénée pouvant porter jusqu'à 3 chlores) (Figure 21).

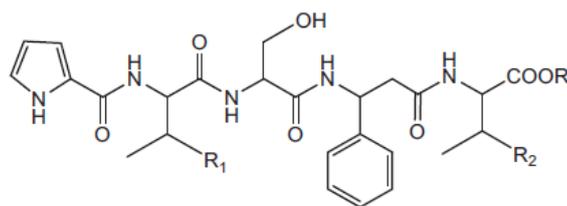


	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	X-Y
1	H	OH	Cl	Cl	H	OH	CH-CH
2	OH	H	Cl	Cl	Cl	OH	CH-CH
3	H	H	Cl	Cl	H	OH	CH-CH
4	H	H	H	H	Cl	OH	C=C
5	OH	H	H	H	Cl	OH	C=C
6	H	H	Cl	H	H	OH	CH-CH
7	H	H	H	H	H	OH	CH-CH
8	H	OH	H	H	Cl	OH	C=C
9	OH	OH	Cl	Cl	H	OH	CH-CH
10	OH	OH	H	H	Cl	OH	C=C
11	H	H	H	Cl	H	OH	CH-CH
12	H	H	H	Cl	H	OH	C=C
13	H	H	Cl	Cl	H	OAc	CH-CH

**Figure 21 : Structure des astines A à H (1 à 8) et K à P (9 à 14). Pour chaque structure, la fonction ou le groupe porté aux différentes positions R1 à R6 est indiqué. Pour les structures de 1 à 13, la nature de la liaison entre les positions notées X et Y sur la proline sont indiquées. Abu : acide amino butyrique. Ava : acide valérique. (Xu *et al.*, 2013)**

La variété de structure des astines repose essentiellement sur des différences au niveau de 3 monomères composant le peptide. La proline en position 1 peut présenter jusqu'à 4 niveaux d'halogénéation différents (de 0 à 3 chlores portés par le monomère). En position 2, les différentes formes d'astines peuvent présenter une *L*-*allo*-thréonine ou un *L*-Abu. Enfin, le monomère en position 5 pourra être une *L*-*allo*-thréonine, un *L*-Abu ou un *L*-Ava. De plus, un niveau d'insaturation différent peut exister entre les carbones  $\gamma$  et  $\delta$ .

Des formes linéaires non halogénées de ces molécules (appelées asterinines) ont également été identifiées, mais considérées comme des artefacts liés à la modification de l'astine par les conditions d'analyses (Liu *et al.*, 2012) (Figure 22).



R	R <sub>1</sub>	R <sub>2</sub>	Name
H	OH	H	asterinin A
Me	OH	H	asterinin B
Me	H	OH	asterinin C
H	H	H	asterinin D
Me	OH	OH	asterinin E
Me	H	H	asterinin F

Figure 22 : Structure des astérinines A à F.

### Activité

La variété de structures des astines entraîne une variété d'activités biologiques. A leur découverte, les astines A et B ont été décrites comme molécules antitumorales. Leur activité a été étudiée sur une lignée de cellules tumorales Sarcoma 180 de souris (Morita *et al.*, 1993). L'astine C présente quant à elle un intérêt en tant qu'immunosuppresseur, permettant l'induction de l'apoptose des lymphocytes T (Shen *et al.*, 2011). En 2013, Xu *et al.* étudient le potentiel antitumoral de 14 formes d'astines (A à H et K à P) sur une variété de lignées cellulaires tumorales, et seules les astines B et C présentent une activité dans les conditions testées (Xu *et al.*, 2013).

Malgré l'intérêt présenté par l'activité antitumorale d'un nombre important de ces molécules, une toxicité secondaire de celles-ci a néanmoins été révélée. En effet, en 2012, Liu *et al* analysent une fraction d'un extrait d'astine présentant une forte hépatotoxicité chez la souris, démontrée par examens histologiques. Dans cette fraction se trouvent 28 pentapeptides (18 cycliques et 10 linéaires) parmi lesquels de nombreuses formes d'astines (de B à I) et de dérivés. L'absence de rapport de la toxicité liée à la consommation d'extraits préparés à partir de racines d'*Aster tataricus* pourrait s'expliquer par la faible dose de peptides consommés, et par le fait que les extraits de plantes préparés dans le cadre de la médecine traditionnelle chinoise concernent généralement des mélanges d'espèces, pouvant entraîner une diminution de la toxicité des astines (Liu *et al.*, 2012).

A ce jour, des activités biologiques telles les activités antibactériennes ou antifongiques n'ont pas été testées.

### 1.3.2 La cyclochlorotine et l'hydroxycyclochlorotine

#### Structure et découverte

La cyclochlorotine (CCT) et l'hydroxycyclochlorotine (HCCT) sont deux pentapeptides cycliques identifiés dans des surnageants de culture de *Penicillium islandicum*. Cette moisissure produit une large variété de toxines (simatoxine, luteoskyrine, islanditoxine) et de pigments (Ueno et Ishikawa, 1969 ; Ghosh *et al.*, 1978).



Figure 23 :A. Hyphe de *Penicillium islandicum* en microscopie optique. B. Culture de *Penicillium islandicum* sur PDA après 12 jours à 25°C.

Ces molécules ont été découvertes après l'identification d'une moisissure contaminant des stocks de riz destiné à la consommation humaine au Japon dans les années 50. Une substance hépatotoxique a alors été isolée de cultures de *P. islandicum* et testée sur des rats, provoquant leur mort rapide (Tatsuno *et al.*, 1955). Dans les années 1970, la structure de la cyclochlorotine est élucidée par cristallographie aux rayons X, révélant un pentapeptide cyclique de séquence  $(Cl_2)Pro_1-Abu_2-Ser_3-Phe_4-Ser_5$  (Ishikawa *et al.*, 1970, Ishii et Ueno, 1973, Yoshioka *et al.*, 1973). La structure de l'hydroxycyclochlorotine a été élucidée lors de sa découverte, révélant une différence du monomère en position 2, l'HCCT portant une thréonine, là où la CCT porte un acide aminobutyrique (Mizutani *et al.*, 2008) (Figure 24).

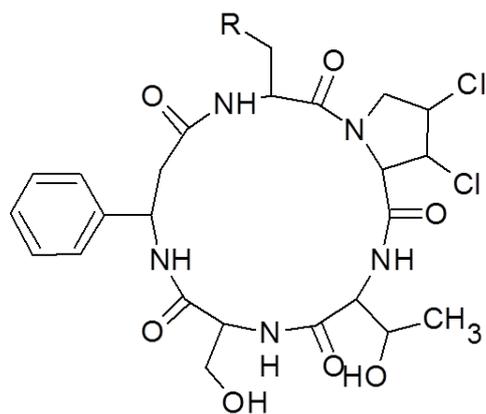


Figure 24 : Structure de la cyclochlorotrypsine (R = H) et de l'hydroxycyclochlorotrypsine (R = OH)

### Activité

L'activité de la cyclochlorotrypsine est connue depuis sa découverte, cette activité ayant été à l'origine des recherches effectuées sur cette molécule. En effet, c'est la toxicité du riz contaminé par le *Penicillium islandicum* produisant la CCT qui a déclenché les recherches sur cette molécule, posant alors un problème de sécurité sanitaire. Les différents tests de toxicité réalisés sur des rongeurs ont conclu à son effet hépatotoxique (Terao *et al.*, 1984). La présence des chlorures associés à la proline semble jouer un rôle important dans la toxicité de la molécule (Ueno, 1992).

### Similarités entre astine, cyclochlorotrypsine et hydroxycyclochlorotrypsine :

La comparaison des structures des deux familles de molécules, à savoir les astines isolées d'*Aster tataricus* d'une part, et la cyclochlorotrypsine et l'hydroxycyclochlorotrypsine issues de *Penicillium islandicum* d'autre part permet de mettre en évidence la similarité de structures de ces deux groupes de peptides. En effet, la différence entre l'astine B et la cyclochlorotrypsine tient à la différence d'un monomère de la chaîne peptidique (Figure 25).

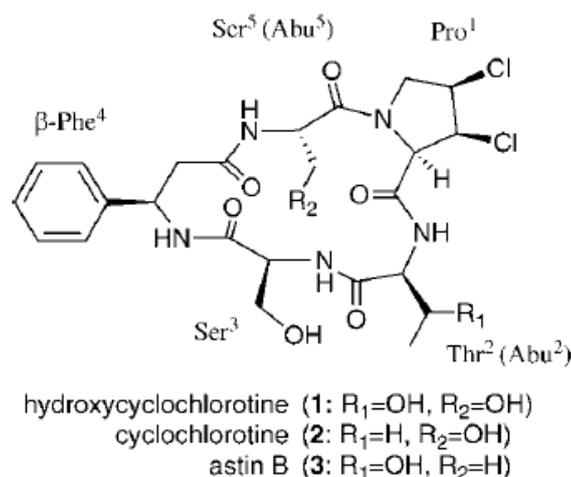


Figure 25 : Structure de l'hydroxycyclochlorotine, de la cyclochlorotine et de l'astine B

Il est intéressant de noter cette forte ressemblance structurale entre deux familles de molécules connues comme étant des produits d'une moisissure d'une part, et d'une plante d'autre part. Cette similarité est à l'origine d'une hypothèse de départ dans les travaux visant à étudier le mécanisme de synthèse de ces différentes molécules

Lorsque les travaux présentés dans ce manuscrit concerneront sans distinction les deux familles de peptides, ceux-ci seront regroupés sous le terme de PentaPeptides Cycliques Chlorés, ou PPCC.

### 1.3.3 Mécanisme de synthèse des PPCC : Un mécanisme NRPS ?

A ce jour, le mécanisme de synthèse des différents PPCC n'a pas été identifié. La CCT et l'HCCT ayant été isolées à partir de cultures pures de *Penicillium islandicum*, celui-ci est probablement responsable de leur production. Les différentes formes d'astines ayant été isolées à partir de racines d'*Aster tataricus*, ont de la même manière été considérées comme des produits de la plante. Toutes ces molécules présentent des particularités structurales peu communes :

- Ce sont des peptides cycliques.
- Ils portent des acides aminés non protéogéniques, tels que l'acide aminobutyrique.
- La proline peut présenter différents niveaux d'halogénéation, avec la présence de 0, 1, 2 ou 3 atomes de chlores.

Ces différents éléments peuvent amener à développer l'hypothèse d'une voie de synthèse non-ribosomique pour ces molécules bien qu'aucune voie de synthèse de type NRPS n'ait jamais été identifiée chez des végétaux supérieurs.

Il est également important de noter que les peptides cycliques comme les PPCC présentés dans ce manuscrit sont considérés comme marqueurs d'infections fongiques des plantes (Jegorov *et al.*, 2006).

Si la proximité des structures des deux familles de molécules a été mise en avant dans la littérature, aucune recherche n'a été réalisée à notre connaissance à ce jour pour déterminer les gènes de production des deux familles de molécules, ou pour analyser la raison de cette proximité structurale.

De même, aucune étude permettant de mettre en évidence un lien entre *Aster tataricus* et *Penicillium islandicum* d'un point de vue écologique n'a été réalisée.

## 2. Situation et objectif des travaux

Les travaux réalisés dans le cadre de cette thèse ont pour objectif d'apporter de nouveaux éléments dans la compréhension du mécanisme de synthèse des deux familles de molécules étudiées, les kurstakines d'une part, et les PPCC d'autre part.

Les deux modèles utilisés ici présentent des différences notables :

- Le type d'organisme producteur : La kurstakine est produite par *Bacillus thuringiensis*, un organisme procaryote, tandis que les PPCC sont produits par l'organisme eucaryote *Penicillium islandicum* pour la cyclochlorotine et l'hydroxycyclochlorotine. Les astines, quant à elles, sont connues comme étant un produit de la plante *Aster tataricus*.
- Le niveau de connaissance global associé à chacune des deux familles de molécules. Ainsi, si les gènes de production de la kurstakine synthétase et certains éléments de la régulation de ces gènes sont connus, les gènes responsables de la production des PPCC n'ont à ce jour pas été identifiés.
- Les activités des deux types de molécules n'ont pas été aussi clairement identifiées. Si l'activité antitumorale des astines et le caractère hépatotoxique des PPCC sont aujourd'hui démontrés par de nombreuses études, il n'existe à ce jour que peu d'informations sur les activités biologiques de la kurstakine.

Les deux modèles présentent néanmoins une problématique similaire : Si les deux familles de molécules étudiées dans ce manuscrit possèdent un fort potentiel dans le cadre d'une valorisation industrielle, aucun procédé de production permettant d'obtenir les différentes molécules en des quantités suffisamment importantes pour valider ou développer ce potentiel n'a été mis au point.

Le but des travaux réalisés dans le cadre de cette thèse est d'apporter de nouveaux éléments de compréhension du mécanisme de production des deux familles de molécules permettant à terme de développer des procédés de production efficaces pour les différents peptides.

Dans le cadre des travaux sur la kurstakine, le niveau de production du lipopeptide par différentes souches sera déterminé afin de choisir la souche sauvage produisant la quantité la plus importante.

L'effet d'une série de mutations et de modifications effectuées dans le génome de la souche sur le niveau de production du lipopeptide sera alors analysé.

Comme présenté dans la synthèse bibliographique, la kurstakine semble répondre à une dynamique de production différente des autres lipopeptides produits par le genre *Bacillus*, présentant une production à des temps très tardifs de la culture dans la littérature. La cinétique de production de la kurstakine sera étudiée en milieu solide et liquide afin de mieux comprendre ce phénomène.

Dans le cadre des travaux sur les PPCC, une recherche des gènes de production de la cyclochlorotine et de l'hydroxycyclochlorotine sera effectuée par une analyse bioinformatique.

Une étude de l'activité enzymatique des domaines A de NRPS identifiés par bioinformatique et potentiellement impliqués dans la production des PPCC sera effectuée *in vitro* et *in vivo*. L'étude *in vitro* sera réalisée par le clonage des domaines A dans des cellules hôtes pour l'expression hétérologue de ces domaines et l'analyse de leur activité. L'étude *in vivo* sera réalisée par des expériences d'alimentation différenciée en acides aminés composant les PPCC ou étant les précurseurs des monomères entrant dans leur composition. Les produits obtenus à l'issue de ces expériences d'alimentation différenciée seront analysés par spectrométrie de masse.

# **CHAPITRE 2**

## **Matériels et méthodes**

# 1. Souches et plasmides utilisés

## 1.1 Souches microbiennes

Les souches de *Bacillus thuringiensis* (Bt) utilisées lors de cette étude ont été fournies par l'équipe MICALIS de l'INRA de Jouy-en-Josas. Ce sont, d'une part des souches sauvages isolées de l'environnement, et d'autre part, des souches modifiées génétiquement pour l'étude de l'effet de différents facteurs sur la production de kurstakine par Bt.

Tableau 3 : Liste des souches de *Bacillus thuringiensis* utilisée dans le cadre de l'étude du mécanisme de production de la kurstakine

Souche <i>B. thuringiensis</i>
<i>Bacillus thuringiensis</i> 407
<i>Bacillus thuringiensis</i> 407 cry-
<i>Bacillus thuringiensis</i> 407 $\Delta$ spo0A
<i>Bacillus thuringiensis</i> HD1
<i>Bacillus thuringiensis</i> HD73
<i>Bacillus thuringiensis</i> HD73 cry-
<i>Bacillus thuringiensis</i> HD73 Cry- [Pcry3A-stab-SD::krs]
<i>Bacillus thuringiensis</i> HD73 Cry- [Pcry3A-stab-SD::krs] / pHT1618T-PxylQkrsE
<i>Bacillus thuringiensis</i> HD73 cry- stabSD spo0A

Les souches bactériennes utilisées dans le cadre de cette étude sont mises en culture dans 5 mL de milieu de culture LB liquide à 37°C, agitées sur un agitateur rotatif à 160 rpm jusqu'à l'obtention d'une DO<sub>600nm</sub> de 1. La culture est alors répartie dans des cryotubes de 1 mL selon les proportions suivantes : 0,3 mL de glycérol pour 0,7 mL de culture. Les cryotubes ainsi obtenus sont directement mis au congélateur à -80°C pour le stockage de longue durée des souches bactériennes.

## 1.2 Plasmides

### pCSC89

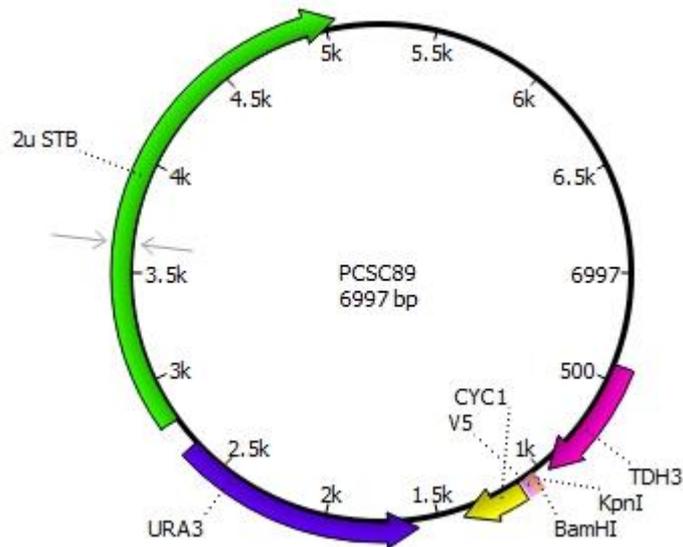


Figure 26 : Carte du plasmide pCSC89 (Légende : 2uSTB : Origine de répllication 2 $\mu$  ; URA3 : gène de l'orotidine 5'-phosphate décarboxylase ; CYC1 : Terminateur ; V5 : Epitope V5 ; TDH3 : Promoteur de la glyceraldehyde 3-phosphate déhydrogénase)

Le plasmide pCSC89 est utilisé pour le clonage chez *Saccharomyces cerevisiae*. Il comprend une séquence codant pour un polyHis à l'extrémité N-terminale de la protéine recombinante, et une séquence codant pour un épitope V5 à l'extrémité C-terminale, ainsi que des sites de restrictions *KpnI* et *BamHI* de part et d'autre du site de clonage. Il comprend également le gène du marqueur d'auxotrophie *ura3*, codant pour l'orotidine 5'-phosphate décarboxylase, enzyme impliquée dans la synthèse des bases pyrimidiques. Enfin, le pCSC89 porte également un site de reconnaissance pour la FLP recombinase, nécessaire à la réalisation de la recombinaison entre le vecteur et l'insert par la souche pour la méthode de gap-repair cloning (Chino *et al.*, 2010)

### pPIC3.5

Le plasmide pPIC3.5 est utilisé pour le clonage chez *Pichia pastoris* et l'expression intracellulaire de protéines recombinantes. Il possède le gène de résistance à l'ampicilline pour la sélection chez l'hôte intermédiaire *Escherichia coli* et le marqueur de sélection HIS4 pour la sélection des transformants chez *P. pastoris*. En effet, les souches utilisées pour le clonage, GS115 et KM71, présentent une mutation au niveau du gène de l'histidinol déhydrogénase *his4*, les empêchant de synthétiser l'histidine. Le plasmide possède des sites de restrictions uniques pour les enzymes *BamHI*, *SnaBI*, *EcoRI*, *AvrII* et *NotI* pour la

réalisation du clonage. L'insertion du produit PCR s'effectue en aval du gène de l'alcool oxydase AOX1, permettant l'induction au méthanol de la transcription de la protéine recombinante.

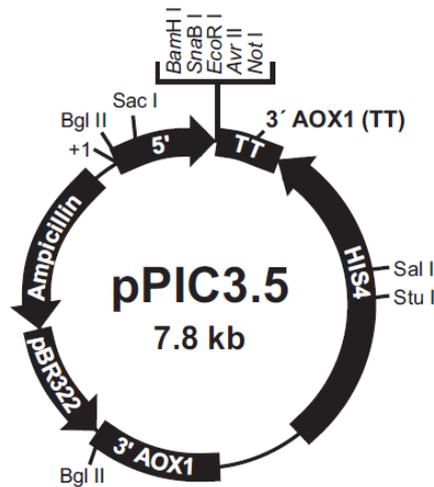


Figure 27 : Carte du plasmide pPIC3.5. Légende : 3'AOX1 : fragment 3' de l'alcool oxydase AOX1 ; HIS4 : gène de l'histidinol déshydrogénase ; TT : Terminateur de transcription d'AOX1 ; 5' : fragment 5' de l'alcool oxydase AOX1 ; Ampicillin : Gène de résistance à l'ampicilline ; pBR322 : origine de réplication.

### 1.3. Milieux de culture

#### *Milieu LB*

Le milieu de culture Luria Bertani (LB) est un milieu complexe classiquement utilisé pour la mise en culture des souches bactériennes. Il est composé de tryptone, 10 g/L ; extrait de levure, 5 g/L ; NaCl, 10 g/L. Les composants sont solubilisés dans 800 mL d'eau déminéralisée, le pH est ensuite ajusté à 7,0 et le volume est complété à 1 L. Ce milieu peut être solidifié par l'addition de d'agar (15 g /L). . Le milieu est stérilisé par autoclavage à 121°C pendant 15 minutes.

#### *Milieu 863*

Le milieu 863 est un milieu utilisé pour la mise en culture des souches de levures utilisées pour les manipulations de biologie moléculaire. La composition du milieu 863 est la suivante : glucose, 10 g/L ; extrait de levure 10 g/L ; peptone de caséine, 10 g/L. Le milieu gélosé 868 peut être préparé en ajoutant de l'agar (15 g/L). Le milieu est stérilisé par autoclavage à 121°C pendant 15 minutes.

### *Milieu Czapek*

Le milieu Czapek est un milieu de culture utilisé pour la culture des moisissures, pour la production et la détection des PPCC. Sa composition est la suivante : saccharose, 30 g/L ; K<sub>2</sub>HPO<sub>4</sub>, 2 g/L ; MgSO<sub>4</sub>, 0,5 g/L ; KCl, 0,5 g/L ; FeSO<sub>4</sub>, 0,01 g/L. Le milieu gélosé Czapek-agar est réalisé en ajoutant 15 g d'agar/L. Le milieu est stérilisé par autoclavage à 121°C pendant 15 minutes.

### Milieux Czapek modifiés

Dans le cadre d'expérience d'alimentation différenciée en acides aminés pour la production de dérivés des PPCC, le milieu Czapek peut être modifié par l'addition de peptone à 1%, et par l'ajout de différents acides aminés. Une série de milieux modifiés est ainsi préparée en ajoutant au Czapek ou au Czapek peptone l'un des acides aminés suivant à la concentration indiquée : thréonine ; 1 g/L, tryptophane, 0,2 g/L ; tyrosine, 0,4 g/L ; glycine, 1 g/L ; alanine, 1 g/L ; valine, 0,92 g/L ; leucine, 1 g/L ; isoleucine, 1 g/L.

### *Milieu AK agar #2*

Le milieu AK agar #2 (Beckton, Dickinson, Etats-Unis) est un milieu acheté prêt à l'emploi utilisé pour la sporulation des souches de *Bacillus*. La composition approximative indiquée par le fournisseur est la suivante : peptone de gélatine, 6 g/L ; peptone de caséine, 4 g/L ; extrait de levure, 3g/L ; extrait de bœuf, 1,5 g/L ; glucose, 1 g/L ; agar, 15 g/L ; sulfate de manganèse II, 0,3 g/L. Le milieu est préparé en solubilisant 30,8 g de milieu dans 1 L d'eau déminéralisée, et est stérilisé par autoclavage à 121°C pendant 15 minutes.

### *Milieu PDA*

Le milieu PDA (Potato Dextrose Agar, Biokar Diagnostics) est le milieu utilisé pour la mise en culture des souches de moisissures et pour les tests d'activité antifongique. La composition du milieu est la suivante : extrait de pomme de terre, 4 g/L ; glucose, 20 g/L ; agar, 15 g/L. Le milieu est préparé en solubilisant 39 g de milieu dans 1 L d'eau déminéralisée et est stérilisé par autoclavage à 121°C pendant 15 minutes.

### *Milieu SD medium-uracil*

Le milieu SD medium-URA (MP Biomedicals) est utilisé pour la sélection des transformants obtenus dans le cadre du clonage de domaines d'adénylation chez la levure. Sa composition est la suivante : yeast nitrogen base (YNB), 1,7 g/L ; (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 g/L ; glucose,

20 g/L, complete supplement mixture sans uracil (CSM-ura), 0,69 g/L. Le milieu est préparé en solubilisant 27,4 g de milieu dans 1 L d'eau déminéralisée, auxquels peuvent être rajoutés 15 g d'agar pour l'obtention d'un milieu gélosé. Il est stérilisé par autoclavage à 121°C pendant 15 minutes.

## **2. Production de kurstakine**

### **2.1 Production en milieu solide**

Les souches de *B. thuringiensis* ont été cultivées sur milieux solide (AK agar ou LB gélosé) pour la production de kurstakines selon le protocole décrit par Hathout *et al.*, 2000. Une pré-culture d'une nuit est réalisée dans du milieu LB liquide à 30°C, agitée à 160 rpm. Les boîtes de milieu gélosé sont alorsensemencées par inondation, et le surplus de pré-culture est éliminé. Les boîtes sont mises à incuber une nuit à 37°C et 3 jours à température ambiante sur pailleasse. Le prélèvement de la culture sur boîte est réalisé après 24, 48, 72 et 120 heures de culture pour le dosage des lipopeptides.

Le tapis bactérien s'étant développé sur les boîtes de milieu gélosé est récupéré au moyen d'un râteau, et repris dans une solution à 1 M KCl/0,5 M NaCl. La solution est homogénéisée au vortex, puis centrifugée à 10 000 g pendant 10 minutes. Le culot est ensuite lavé 3 fois par un volume équivalent d'eau distillée. L'extraction est réalisée sur ces culots par ajout de solvant ACN/H<sub>2</sub>O 90/10 (V/V), suivi d'une homogénéisation au vortex. La suspension est ensuite centrifugée à 3 000 g pendant 10 minutes, afin de séparer le surnageant, contenant les composants récupérés de la surface des spores, et le culot cellulaire. Le surnageant est ensuite directement traité par extraction en phase solide (SPE) pour l'extraction des kurstakines.

### **2.2 Production en milieu liquide**

#### *2.2.1 Production en fioles Erlenmeyer*

Pour les cultures bactériennes, une culture de réveil est réalisée dans 10 mL de milieu LB dans un tube à essai à partir d'un cryotube de la souche conservée à -80°C.

Les cultures liquides de Bt sont réalisées à partir des cryotubes préparés précédemment. Une pré-culture est réalisée dans 100 mL de milieu de culture liquide correspondant au milieu utilisé pour la culture. La croissance de la pré-culture est réalisée dans les mêmes conditions de température et d'agitation que la culture pour laquelle celle-ci a été préparée. Lorsque la

pré-culture atteint une  $DO_{600nm} \approx 1$ , elle est utilisée pour ensemercer la culture pour atteindre une  $DO_{600nm}$  initiale égale à 0,2.

### 2.2.2 Production en fermenteur

La production de kurstakine en fermenteur est réalisée dans un fermenteur d'une capacité de 5L, avec 3L de milieu LB. Le fermenteur est incubé avec une pré-culture d'une nuit réalisée en milieu LB à 30°C avec une agitation de 160rpm. Le milieu LB du fermenteur est inoculé avec la pré-culture pour atteindre une  $DO_{600nm}$  de 0,2. La température de culture est de 30°C et la pression partielle en oxygène est maintenue à 20% au moyen de l'agitation. Le pH est maintenu à 7,0 par l'apport de  $H_3PO_4$  1M et de NaOH 3M.

## 2.3 Purification

La purification des kurstakines est effectuée par SPE sur colonnes C8 adaptées sur un extracteur sous vide Alltech Vacuum 12-port Manifold, au moyen d'une pompe Savant Gel Pump GP100. Les colonnes SPE C8 (Bond-Elut Jr C8 1000Mg) sont tout d'abord lavées et conditionnées par le passage de 20 mL d'ACN /0,1% TFA, puis 20 mL d' $H_2O$ /0,1% TFA. L'échantillon à analyser (5 mL d'extrait cellulaire ou de surnageant de culture) est alors déposé sur la colonne et entièrement passé à travers la colonne. Un premier lavage est effectué avec 10 mL d' $H_2O$ /0,1%TFA afin d'éliminer les sels et impuretés hydrophiles contenus dans l'échantillon. Un deuxième lavage est effectué par le passage de 10 mL d'ACN/ $H_2O$ /TFA (30/70/0,1 ; v/v/v). Les lipopeptides sont élués par le passage de 5 mL d'ACN/ $H_2O$ /TFA (90/10/0,1 ; v/v/v). Les 5 ml d'échantillon ainsi collectés sont séchés au moyen d'un concentrateur rotatif sous vide. Les culots sont repris dans 300  $\mu$ L d'ACN/ $H_2O$ /TFA (30/70/0,1 ; v/v/v) afin d'être analysés en HPLC.

La mise au point de la purification de kurstakine par FPLC a été réalisée afin de permettre la purification du lipopeptide à partir de volumes importants d'extraits cellulaires de *Bacillus thuringiensis* ne pouvant être réalisés par colonnes SPE de faible volume.

La purification a été réalisée au moyen d'une colonne en verre de 15 cm de hauteur pour 2 cm de diamètre rempli au moyen de phase BONDESIL-C8, 40 $\mu$ m (Agilent Technologies). Les solvants sont pour la voie A de l'acétonitrile additionné de TFA à 0,1% (v/v) et pour la voie B de l'eau ultrapure contenant 0,1% de TFA (v/v). Un volume de 5 à 10 ml d'extrait de kurstakine concentré est injecté dans la colonne. La séparation des kurstakines est réalisée au moyen d'un programme en 4 phases basé sur le programme utilisé en HPLC : 25 minutes à

100% d'eau pour l'élution des contaminants hydrophiles, 20 minutes à 30% d'acétonitrile pour l'élution d'une partie des contaminants présentant une affinité avec ce solvant, 20 minutes de passage de solvant à 70% d'acétonitrile pour l'élution des kurstakines, et enfin 20 minutes à 100% d'acétonitrile pour le nettoyage de la colonne. La récolte de la kurstakine est réalisée au moyen d'un collecteur automatique pendant la phase d'élution à 70% d'acétonitrile.

## **2.4 Détection et quantification**

L'analyse est réalisée au moyen d'une colonne Kinetex C8 (ref) sur un appareillage Waters (ref). La séparation des kurstakines est effectuée par un gradient linéaire d'acétonitrile additionné de 0,1% de TFA (solvant A) de 30 à 70% sur 25 minutes. La colonne HPLC subit ensuite un lavage pendant 5 minutes avec 100% de solvant A, puis 5 minutes de rééquilibrage à 30% du solvant A pour le passage de l'échantillon suivant. La détection des lipopeptides est effectuée par lecture de l'absorbance des liaisons amides à 214 nm. Le temps de rétention des pics obtenus et leurs dérivées secondes sont analysés au moyen du logiciel EMPOWER 2 Software. La quantification de la kurstakine est réalisée par comparaison avec l'aire de pics obtenus pour un standard de surfactine à 1 g/L analysé dans les mêmes conditions.

La standardisation des dosages de lipopeptides est réalisée en rapportant les quantités de lipopeptides mesurées à la masse sèche biologique à partir de laquelle la kurstakine a été extraite. Dans le cadre des analyses de production en milieu solide, les cellules sont récupérées après l'extraction sur SPE par reprise dans 5 mL d'eau déminéralisée et la suspension obtenue est déposée dans une coupelle en aluminium préalablement pesée et mise à sécher dans une étuve à 95°C pendant 48 h pour l'élimination totale de l'eau de l'échantillon. Après évaporation, la coupelle est à nouveau pesée pour déterminer la masse sèche de l'échantillon.

## **3. Production de PPCC**

Les PPCC sont produits en milieu liquide Czapek ou Czapek modifié par *P. islandicum*. Une petite quantité de mycélium est récupérée à partir d'une culture en milieu PDA de la moisissure et reprise dans 100 ml de milieu liquide Czapek ou Czapek modifié, puis mise en culture à 20°C sans agitation pendant 21 jours.

## 4. Techniques analytiques

### 4.1 HPLC

Les analyses HPLC sont réalisées au moyen d'un appareillage Waters (Online Degaser, 717 Autosampler, 660S Controller, 626 Pump, 2996 PhotoDiodeArray, Waters Corporation, Milford, MA, U.S.A.)

### 4.2 Spectrométrie de masse

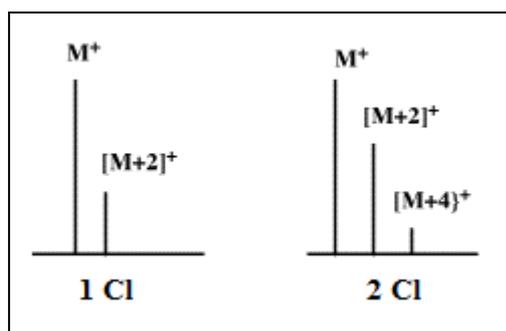
La recherche des différents peptides synthétisés par *B. thuringiensis* est réalisée par spectrométrie de masse à partir de colonies bactériennes développées sur AK agar ou milieu LB gélosé pendant 24 et 48h. Une colonie est en mise en suspension dans 25  $\mu\text{L}$  de matrice HCCA (Acide  $\alpha$ -cyano-4-hydroxycinnamique). Dans le cadre de l'étude de moisissure, quelques mg de mycélium sont récoltés à partir d'une culture sur milieu gélosé.

La suspension est homogénéisée au vortex pendant 10 secondes, centrifugée rapidement (short spin) à 5000 rpm pendant 5 secondes et le surnageant est déposé sur plaque pour analyse. L'acquisition des spectres de masse est réalisée au moyen d'un instrument MALDI-TOF-TOF UltraFlex II (Bruker Daltonics, Brême, Allemagne), équipé avec un laser Smartbeam, en mode réflectron positif. Un millier de spectres MALDI-MS ont été acquis à chaque position dans un champ de 500 à 2000 Da en utilisant une fréquence de laser de 20 Hz.

L'analyse des surnageants de culture et d'échantillons liquides est réalisée en mélangeant 10  $\mu\text{L}$  d'échantillon avec 10  $\mu\text{L}$  de matrice HCCA.

Afin d'identifier les molécules recherchées, les pics correspondants aux adduits  $\text{H}^+$ ,  $\text{Na}^+$  et  $\text{K}^+$  de celles-ci sont recherchés.

Dans le cadre des travaux sur les PPCC, l'analyse du profil des massifs isotopiques spécifiques est également réalisée pour la recherche de molécules halogénées.



**Figure 28 : Profil du massif isotopique observé en spectrométrie de masse  
d'une molécule contenant 1 ou 2 atomes de Cl**

En effet, la présence d'atomes de chlore dans la formule moléculaire va influencer sur le profil du massif isotopique de la molécule correspondante, en raison de l'existence des deux isotopes du chlore  $^{35}\text{Cl}$  (75%) et  $^{37}\text{Cl}$  (25%). L'existence de ces isotopes provoque ainsi la présence de pics caractéristiques en M+2 (soit M, représentant la masse de la molécule identifiée, additionné de 2 Daltons), correspondant à 32% de l'intensité du pic M pour une molécule contenant un atome de chlore, et 64% si deux atomes de chlores sont présents. Dans ce dernier cas, un pic à M+4 représentant 10% de l'intensité du pic M peut également être observé.

## 5. Techniques de biologie moléculaire

### 5.1 Extraction de l'ADN de moisissure

La technique d'extraction d'ADN de moisissure est réalisée selon un protocole modifié de Moller *et al.* 1992. La souche est d'abord cultivée sur milieu PDA à 20°C pendant 5 jours. L'extraction est réalisée en 4 étapes à partir de 100 mg de mycélium récolté à partir de culture en milieu liquide ou solide :

-Digestion du mycélium : 100 mg de mycélium sont broyés dans 500  $\mu\text{L}$  de tampon d'extraction (100 mM TrisHCl, 10 mM EDTA, 100  $\mu\text{g/ml}$  de protéinase K, 1 % 2-mercaptoéthanol, 2 % SDS, qsp 100 mL d'eau distillée) et incubés à 60°C pendant 60 minutes. 140  $\mu\text{L}$  de NaCl 5 M et 65  $\mu\text{L}$  de bromure d'hexadécyltriméthylammonium (CTAB) à 10% sont ajoutés. Le mélange est incubé à 65°C pendant 10 minutes.

-Dénaturation des protéines : Un volume de chloroforme est ajouté à la préparation précédente, et le tout est mis à incuber à 0°C pendant 30 minutes. Une centrifugation à 14 000 rpm pendant 10 minutes à 4°C est effectuée. La phase supérieure est récupérée et additionnée

d'un volume d'acétate d'ammonium 4M. Le mélange est incubé sur de la glace pendant 60 min, puis centrifugé à 14 000 rpm pendant 10 min à 4°C.

-Précipitation de l'ADN : le surnageant est transféré dans un nouveau tube auquel est ajouté de la RNase A à 10 mg/mL (concentration finale 0,02 µg/µL) et 0,5 volume d'isopropanol. Le mélange est centrifugé à 14 000 rpm pendant 10 minutes à 4°C.

-Lavage et récupération de l'ADN : le surnageant est éliminé, et le culot est lavé 3 fois au moyen d'éthanol à 70 %. Après élimination de l'éthanol, le culot est séché à 37°C jusqu'à évaporation complète du solvant. Le culot est enfin repris dans 20 µL d'eau déminéralisée stérile et réhydraté une nuit à 4°C.

La qualité de l'extraction est évaluée en déposant et en faisant migrer 3 µL de l'extraction sur gel d'agarose à 1 %.

## 5.2 PCR

La technique de réaction en chaîne par polymérase (PCR) est réalisée au moyen du mélange Master Mix (Fermentas GMBH, St Leon, Germany) contenant l'enzyme (ADN polymérase), un tampon, des dNTP et du MgCl<sub>2</sub>. A ce mélange est ajouté l'ADN matriciel, les amorces, et de l'eau ultrapure stérile. Pour les PCR réalisées dans le cadre du clonage pour la production de protéines recombinantes, une polymérase possédant une fidélité très importante était requise. Le mélange réactionnel retenu est le Phusion High-Fidelity PCR Master Mix 2X (Thermo Scientific, USA), possédant un taux d'erreur de 0,11 %. Le mélange final utilisé est le suivant :

Phusion High-Fidelity PCR Master Mix 2X	25 µL
Amorce-F	1,5 µL
Amorce-R	1,5 µL
ADN matriciel	x µL
H <sub>2</sub> O	qsp 50 µL

Le mélange est réalisé dans un tube PCR de 200 µL, et est placé dans un thermocycleur Eppendorf Mastercycler Gradient.

Le cycle de température utilisé varie en fonction des amorces utilisées. Les temps des différentes étapes sont les suivants :

Température	Temps	Nb de cycles
98°C	30 s	1
98°C	30 s	
50 à 60°C	30 s	30
72°C	90 s	
72°C	10 min	1

A l'issue de la PCR, les échantillons sont stockés à 4°C pour une utilisation ultérieure à court terme ou à -20°C si le stockage doit excéder 24 heures. 5 µl des échantillons sont également déposés sur gel d'agarose pour migration pour s'assurer de la qualité de la PCR réalisée.

### 5.3 Digestions enzymatiques et Gel d'agarose

Les digestions enzymatiques sont réalisées au moyen d'enzymes de restriction « Fast Digest » (Thermo Scientific, USA), permettant une réaction rapide (5 minutes d'incubation) et l'utilisation de plusieurs enzymes de restriction avec un seul tampon universel pour toutes les enzymes Fast Digest. Le mélange réactionnel pour la digestion enzymatique est le suivant :

Tampon Fast Digest 10X	2 µL
ADN	2 µL (jusqu'à 1 µg)
Enzyme Fast Digest	1 µL
Eau	QSP 20 µL

Le mélange est ensuite incubé à 37°C au thermocycleur. En cas de besoin, les enzymes de restrictions sont inactivées à 65 ou 80°C en fonction de l'enzyme.

Les ADN et plasmides extraits, ainsi que les produits PCR obtenus sont vérifiés au moyen d'une migration sur gel d'agarose. Le gel est préparé en dissolvant l'agarose dans du tampon TBE 0,5X, préparé à partir de TBE5X (Tris 445 mM; acide borique 445 mM; EDTA 10 mM). La quantité d'agarose est ajustée en fonction du volume de gel pour obtenir un gel à 1% d'agarose, et le tout est dissout par chauffage quelques minutes au four à micro-ondes. Le gel est alors coulé dans un support prévu à cet effet comprenant un peigne destiné à la création de puits dans le gel. Après solidification, le gel est inséré dans une cuve à électrophorèse remplie de tampon TBE 0,5X. 5µL des échantillons à analyser sont mélangés à 1µL de tampon de charge et déposés dans les puits du gel. Afin de s'assurer de la taille des fragments obtenus, 5 µL d'un marqueur de poids moléculaire O'Gene Ruler DNA Ladder Mix (Thermo Scientific, USA) est déposé dans un des puits du même gel.

La migration est effectuée en faisant passer dans la cuve un courant de 100 V jusqu'à ce que le front de migration arrive à 1 cm du bord du gel. Le gel est alors retiré de la cuve et révélé aux UV.

## 5.4 Clonage gap-repair cloning

Cette technique de clonage permet l'insertion d'un morceau d'ADN à analyser dans un plasmide sans avoir recours à un clonage *in vitro*. En effet, le principe général est d'insérer dans une cellule un vecteur et un insert tous deux linéarisés et présentant des zones homologues à leurs extrémités. La souche effectue alors la jonction entre les deux morceaux d'ADN par recombinaison homologue.

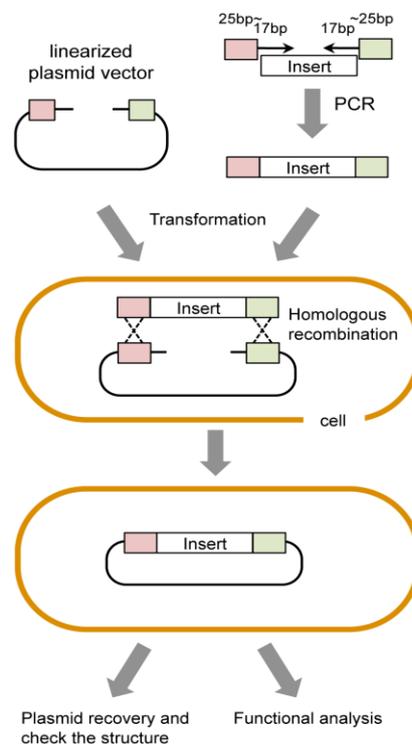


Figure 29 : Schéma de la technique de clonage gap-repair cloning (Chino *et al.*, 2010)

Pour cela, le vecteur et le produit PCR à insérer sont linéarisés au moyen de deux mêmes enzymes de restrictions, possédant des sites de restriction incompatibles entre eux. 10 mL de culture de levure sont préparés dans du milieu 863 et mis à incuber à 25°C, sous agitation à 160 rpm jusqu'à l'obtention d'une  $DO_{660nm}$  de 0,6. La culture est alors centrifugée à 6000 rpm pendant 5 minutes, et le surnageant est éliminé. Le culot est mis en suspension dans 1 mL d'eau stérile, et soumis à une nouvelle centrifugation à 6000 rpm pendant 1 minute. Après élimination du surnageant, le culot est repris par 1 mL d'acétate de lithium à 0,1 M puis soumis à une nouvelle centrifugation à 6000 rpm pendant 1 minute. Le surnageant est éliminé,

et le culot est repris dans 100 µL d'acétate de lithium à 0,1 M, puis incubé une heure à 30°C. Le mélange d'ADN (insert + vecteur linéarisé) est alors ajouté, ainsi que 5 µL d'une solution d'ADN de sperme de hareng à 10 µg/µL (Herring Sperm DNA, Promega) servant d'ADN entraîneur (« DNA carrier »). Le mélange est incubé à 30°C pendant 30 minutes puis 700 µL d'acétate de lithium 0,1 M/ PEG 4000 40 % sont ajoutés au mélange qui est homogénéisé délicatement par retournement des tubes. Le tout est incubé à 30°C pendant une heure, puis placé à 42°C pendant 15 minutes pour réaliser un choc thermique permettant la transformation des cellules. Les cellules sont ensuite centrifugées à 3000 rpm pendant 5 minutes, et le surnageant est éliminé. Le culot est lavé avec de l'eau stérile, et est ensuite repris dans 200 µL d'eau stérile, pour être enfin étalé sur milieu SD medium-uracil agar pour la sélection des transformants.

## 6. Tests d'activité biologique

### 6.1 Souches cibles d'activité antibiotique

Afin d'évaluer le potentiel antifongique de la kurstakine, 4 souches de moisissures phytopathogènes ont été utilisées dans le cadre de tests d'activité. Ces souches ont été fournies par Monica Höfte (Laboratoire de phytopathologie, Département de protection des cultures de l'Université de Gand, Belgique).

Tableau 4 : Liste des souches de moisissures utilisées pour les test d'activités antifongiques

Moisissure	Maladie
<i>Botrytis cinerea</i> R16	Agent de la pourriture grise du raisin
<i>Fusarium oxysporum</i> f. sp. <i>Congutinans</i>	Agent de la fusariose chez les céréales
<i>Rhizoctonia solani</i> S010-1	Agent de la pourriture basale de la laitue
<i>Galactomyces geotrichum</i> MUCL 28959	Agent de la pourriture aigre des fruits

Des tests d'activités antibiotiques ont également été réalisés avec deux souches de collections ATCC : *Escherichia coli* ATCC25922 et *Candida albicans* ATCC10231

## **6.2 Test d'activité antifongique.**

L'activité antifongique des extraits de *Bacillus thuringiensis* enrichis en kurstakine est évaluée au moyen d'un test d'inhibition de croissance de différentes souches de moisissures et de levures sur boîte en milieu solide. Les différentes moisissures à tester sont d'abord cultivées en déposant sur un milieu PDA gélosé un carré de gélose colonisée par ces souches (une souche /boîte de Pétri). Après inoculation les différentes boîtes sont mises à incuber à 25°C jusqu'à la colonisation complète des boîtes. Un carré de gélose de chacune de ces souches sert à coloniser les boîtes utilisées pour le test antifongique. Celui-ci est réalisé en boîtes de Pétri contenant précisément 20ml de PDA. Des carrés de gélose colonisée sont découpés à partir des cultures précédentes et déposés à l'envers au centre des boîtes. L'incubation s'effectue à 25°C jusqu'à l'obtention d'un disque de culture d'1 cm de diamètre. Des disques imprégnés d'extraits purifiés de kurstakines correspondant à 25, 50 et 100µg de lipopeptides sont alors déposés à 1cm du bord de la boîte. Un témoin négatif est réalisé en déposant un disque imprégné du solvant dans lequel ont été repris les extraits de kurstakine (ACN/H<sub>2</sub>O 30/70 (v/v)). Les boîtes sont incubées à 25°C jusqu'à l'envahissement total de la boîte par la souche de moisissure testée. L'effet antifongique de la kurstakine est observé par l'inhibition de la croissance de chaque souche autour des disques imprégnés en lipopeptide.

## **6.3 Test d'activité antibactérienne**

Les tests d'activité antibactérienne sont réalisés par détermination de la CMI sur microplaque 96 puits à fond rond. Un dépôt de 100µL d'échantillon à tester, d'antibiotique témoin ou de milieu de culture est réalisé de la façon suivante : un gradient de concentration de la molécule à tester est réalisé (de 0,4 à 200 µg/mL) dans une série de puits en triple. Une série de dilution d'un antibiotique témoin (témoin positif d'activité) est réalisée dans une série de puits en triple. Enfin, du milieu de culture est déposé dans une série de puits en double (témoin négatif d'activité).

Quelques colonies des souches bactériennes à tester sont mises à incuber dans 10mL de milieu Mueller Hinton à 37°C sous agitation jusqu'à l'obtention d'une DO<sub>600nm</sub> de 1. Les cultures sont alors diluées au dixième pour une DO de 0,1 correspondant à 1.10<sup>8</sup>UFC/mL. Une dilution au 100<sup>ème</sup> de cette première dilution est effectuée pour obtenir une suspension bactérienne à 1.10<sup>6</sup>UFC/mL utilisée pour ensemercer les plaques 96 puits (100µL par puits).

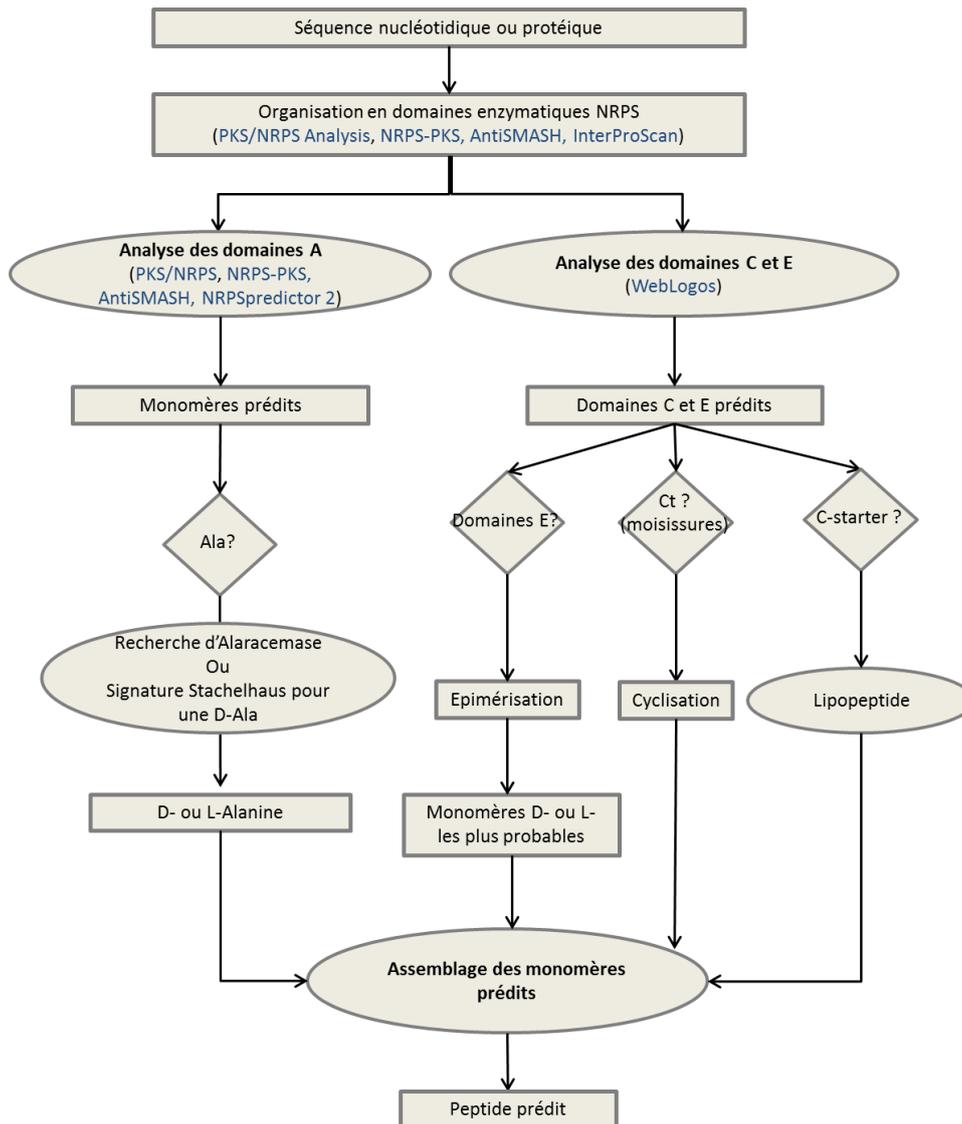
La concentration bactérienne obtenue est validée par la réalisation de dilution successives jusqu'à une concentration de  $1.10^2$ UFC/mL étalées sur milieu de Mueller-Hinton gélosé et mis à incuber 1 nuit à 37°C pour dénombrement.

Une fois les puits incubés avec la suspension bactérienne, les plaques sont mises à incuber à 37°C pendant une nuit. La lecture des plaques est réalisée à l'œil nu par l'observation de la présence ou non d'un trouble dans chaque puits, révélant la croissance bactérienne.

## **7. Analyse *in silico***

### **7.1 Florine : Workflow d'identification de NRPs à partir de séquences génomiques**

La totalité des outils et des méthodes de bioinformatique utilisées dans le cadre des travaux présentés dans ce manuscrit s'inscrit dans un workflow de bioinformatique, permettant d'optimiser l'utilisation des outils de bioinformatique génériques ou spécifiquement dédiés à l'analyse des NRPS pour une meilleure prédiction des produits issus de synthétases détectées dans les génomes et séquences ADN étudiées. Ce workflow permet donc, à partir de séquences ADN ou de protéines, de déterminer le potentiel de production NRPs d'un organisme étudié. La composition en acides aminés constituant le peptide formé par une synthétase identifiée est prédite, et les particularités de structures de ce peptide sont déterminées plus précisément, comme sa cyclisation, l'incorporation probable d'un acide gras (aboutissant à la production d'un lipopeptide), ou encore la présence d'acides aminés sous forme D-.



**Figure 30 : Workflow de bioinformatique utilisé dans le cadre de ces travaux. Les boîtes rectangulaires représentent les données, les ovales représentent les analyses réalisées et les losanges contiennent les questions pouvant amener à l'utilisation de voies d'analyse différentes en fonction du résultat. Les outils et logiciels utilisés sont inscrit en bleu.**

Pour chaque séquence analysée, l'identification des domaines enzymatiques permet alors, pour chacun d'entre eux de déterminer quels outils doivent être utilisés afin d'identifier leur rôle précis.

L'identification d'un domaine d'adénylation orientera donc par exemple vers l'utilisation de logiciels spécifiquement dédiés à l'analyse de ce type de domaines, notamment pour la réalisation d'une prédiction du monomère activé, comme NRPS Predictor 2.

L'identification d'un domaine de condensation, en revanche, orientera vers l'utilisation de logiciels permettant d'identifier le sous-type de domaine C auquel celui-ci appartient, par l'emploi de logiciels permettant cette classification tel NaPDoS, ou l'identification de

séquences consensus spécifiques à un sous-type de domaine par la réalisation d'alignements et de Weblogos.

Cette suite d'outils s'inscrit dans un workflow appelé Florine, définissant les logiciels généraux de bioinformatiques ou spécifiquement dédiés à l'analyse des NRPS quel que soit le type de microorganisme étudié.

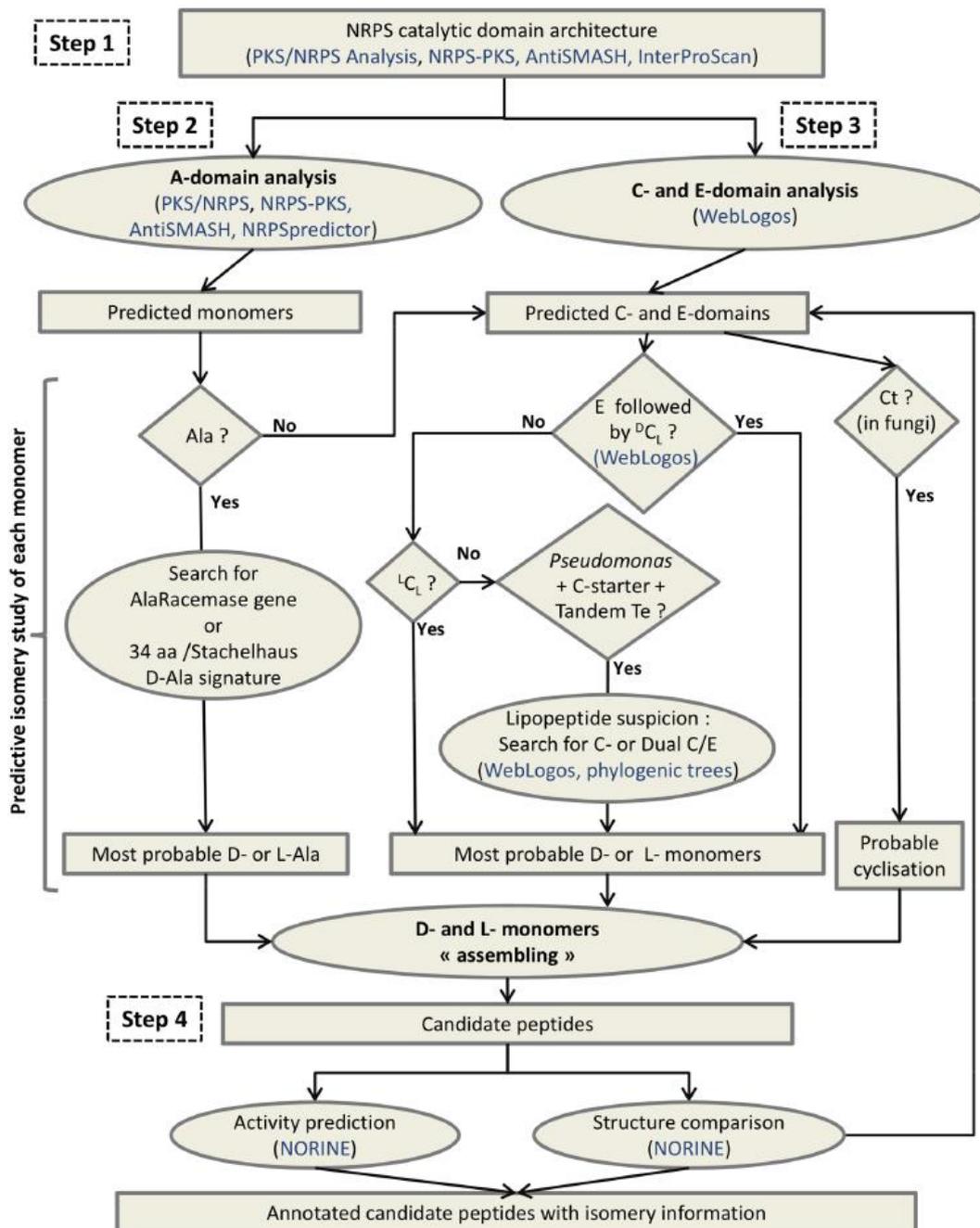


Figure 31 : Totalité du workflow Florine. Les boîtes rectangulaires représentent les données, les ovales représentent les analyses réalisées et les losanges contiennent les questions pouvant amener à l'utilisation de voies d'analyse différentes en fonction du résultat. Les outils et logiciels utilisés sont inscrits en bleu.

Les différents outils utilisés dans ce Workflow, sont présentés dans les paragraphes suivants. En premier lieu sont présentés les outils utilisés pour les analyses globales de bioinformatique (identification de gènes, de signatures spécifiques à des activités enzymatiques particulières, alignements de séquences...), puis sont présentés les outils utilisés spécifiquement pour l'études des gènes et protéines impliqués dans la synthèse non-ribosomique.

## **7.2 Outils généraux d'analyse bioinformatique**

### **NCBI**

Le portail NCBI (National Center for Biotechnology Information) regroupe diverses bases de données et outils liés à l'étude des génomes, gènes et protéines du vivant. On y retrouve ainsi des bases de données de séquences génomiques et protéiques, telles GenBank, comprenant les séquences ADN publiquement disponibles, ou RefSeq, comprenant des séquences annotées et les transcrits et protéines correspondantes. Le portail NCBI représente donc un outil de choix pour la récupération de séquences nucléotidiques ou protéiques au format recherché pour leur analyse au moyen d'outils bioinformatiques. La recherche de séquences au sein des bases de données du NCBI est effectuée par mots-clés, et peut être affinée en déterminant un organisme ou groupe d'organismes portant en leur génome la séquence ou le gène recherché, ou au moyen de fonctions logiques (AND, OR, AND NOT par exemple). Les séquences identifiées sont récupérées au format FASTA pour leur utilisation dans le cadre d'études de bioinformatiques avec différents outils présentés ci-dessous.

### **BLAST : Outil d'alignement de séquences nucléotidique ou protéique**

La recherche d'homologie de séquences est réalisée au moyen du logiciel BLAST du portail NCBI (<http://www.ncbi.nlm.nih.gov/>). L'outil BLAST permet de rechercher dans les bases de données du NCBI des séquences nucléotidiques ou protéiques proches de celle analysée, ou de déterminer l'homologie entre deux séquences ou plus.

La recherche d'homologie de séquence est donc réalisée à partir d'une séquence nucléotidique ou protéique au format FASTA utilisée en tant que requête ou « query ». La recherche d'homologie d'une ou plusieurs autres séquences vis-à-vis de cette requête est réalisée en entrant ces autres séquences en tant que « subject » au format FASTA également.

### **MUSCLE : outil pour la construction d'alignements multiples**

Les alignements multiples de séquences pour l'étude phylogénétique des séquences sont réalisés au moyen de MUSCLE utilisant l'algorithme d'alignement UPGMA (Unweighted Pair Group Method with Arithmetic mean), destiné à la création d'arbre phylogénétique. L'utilisation de MUSCLE est réalisée au sein du logiciel MEGA6, regroupant plusieurs outils pour l'alignement de séquences ADN ou protéiques et la construction d'arbres phylogénétiques. L'alignement obtenu est exporté aux formats MEGA et FASTA pour l'usage ultérieur avec d'autres logiciels d'analyses.

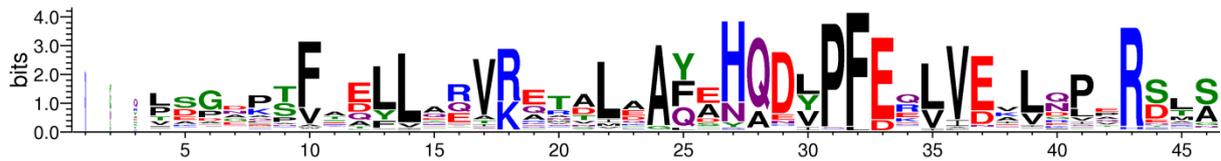
### **Weblogo : Outil pour l'identification de positions conservées.**

Les weblogos sont réalisés au moyen du logiciel Weblogo 3 à partir de fichiers FASTA d'alignements réalisés précédemment. La conservation d'une position de la séquence en acides aminés est représentée par la hauteur de la lettre représentant l'acide aminé en question selon le code international de correspondance à une lettre (Figure 32). La hauteur des lettres est mesurée en bits selon la formule de calcul du Rseq, représentant la conservation de la séquence à une position comme étant la différence entre l'entropie maximale (Smax) et l'entropie observée (Sobs) (Schneider *et al.*, 1990) :

$$R_{seq} = S_{max} - S_{obs} = \log_2 N - \left( - \sum_{n=1}^N p_n \log_2 p_n \right)$$

**Figure 32 : Calcul du Rseq utilisé pour la représentation visuelle de la conservation d'une position au sein d'une séquence.**

où  $P_n$  représente la fréquence observée du symbole de l'acide aminé  $n$  à une position précise de la séquence et  $N$  est le nombre de symboles différents utilisés pour le type de séquence utilisé (soit 4 pour les séquences nucléotidiques, et 20 pour les protéines). Le format de sortie est choisi pour représenter les acides aminés en couleur correspondant à leurs propriétés physico-chimiques. Ainsi, les acides aminés polaires (G, S, T, Y, C) sont représentés en vert, les basiques (K, R, H) en bleu, les acides (D, E) en rouge, les neutres (Q, N) en violet et les hydrophobes (A, V, L, I, P, W, F, M) en noir.



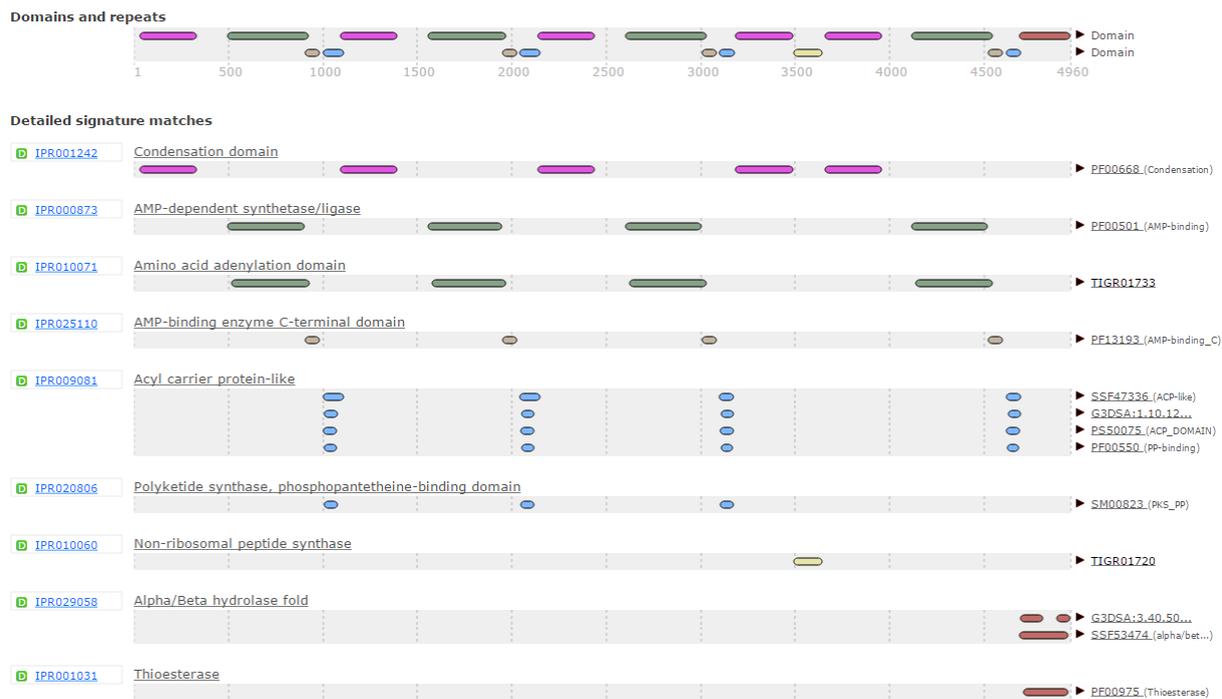
**Figure 33 : Exemple de weblogo réalisé au moyen du logiciel Weblogo 3.0. Les nombres de l'axe horizontal représentent la position au sein de l'alignement de séquences. L'axe vertical représente la conservation à chaque position, mesurée en bits. La couleur des lettres est basée sur la nature physico-chimique de l'acide aminé représenté.**

### **MEGA6 : outil pour construire des arbres phylogénétiques**

Les arbres phylogénétiques sont réalisés au moyen du logiciel MEGA 6 en utilisant les alignements réalisés au moyen de MUSCLE et enregistrés au format MEGA. Les arbres sont réalisés au moyen de la méthode neighbor-joining permettant d'élaborer un arbre représentant la distance génétique entre les séquences analysées. Les paramètres par défaut de la réalisation des arbres selon la méthode neighbour-joining sont utilisés, à savoir l'utilisation de la loi de Poisson comme méthode et le traitement des gaps dans les alignements utilisés par la délétion totale de ces derniers.

### **INTERPROSCAN : outil pour déterminer les domaines enzymatiques**

InterProScan 5 permet l'analyse de protéines par l'utilisation de différentes méthodes de reconnaissance de signatures spécifiques. Les séquences protéiques à analyser sont entrées au format FASTA. Les résultats contenant les signatures détectées avec les scores associés aux hits ayant permis l'identification des signatures peuvent être téléchargés au format XML contenant également une représentation graphique des signatures identifiées dans la séquence analysée.



**Figure 34 :** Exemple de résultat fourni par le logiciel InterProScan. Le premier résultat présente une organisation globale des signatures identifiées dans la séquence protéique. Les lignes suivantes présentent leur répartition dans la séquence par type de signature (en fonction de l'identifiant InterProScan IPR).

## 7.3 Outils spécifiques à l'analyse de NRPS

### NaPDoS : outil pour déterminer les sous-types des domaines C

NaPDoS (Natural Product Domain Seeker) permet la détection et l'analyse de gènes impliqués dans la production de peptides non-ribosomiques, notamment par la recherche des domaines de condensation. L'analyse est réalisée à partir de séquences nucléotidiques ou protéiques. Après identification des domaines de condensation et du sous-type auquel ceux-ci appartiennent, les séquences composant ces domaines peuvent être spécifiquement récupérées au format FASTA, pour leur intégration à un alignement.

### NRPS-PKS : outil pour déterminer les modules et domaines NRPS

L'identification de domaines enzymatiques NRPS est effectuée au moyen du logiciel NRPS-PKS à partir d'une séquence protéique d'une synthétase au format FASTA. Le résultat obtenu est une représentation graphique des domaines enzymatiques sous formes de boîtes de couleur (chaque module étant présenté d'une couleur différente). La séquence composant spécifiquement un domaine peut être récupérée en cliquant sur celui-ci. Une prédiction du monomère activé par les domaines d'adénylation est également réalisée par l'alignement des 10 acides aminés composant le site actif avec ceux de domaines A dont le substrat est connu.

### **PKS-NRPS Analysis : outil pour déterminer les domaines NRPS**

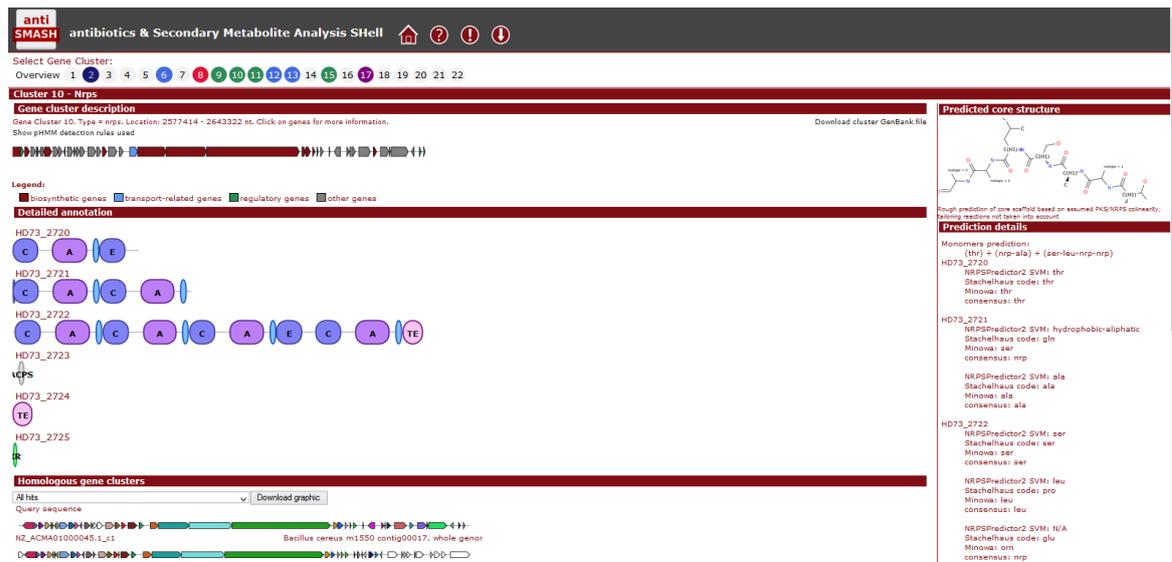
Le logiciel PKS-NRPS Analysis permet l'identification des domaines enzymatiques NRPS à partir de la séquence protéique d'une synthétase. Le résultat obtenu est une représentation graphique des domaines sous formes de disques de couleurs (chaque couleur correspondant à un type de domaine enzymatique). L'organisation en modules de la synthétase n'est pas représentée. Une prédiction du monomère activé par chaque domaine d'adénylation est également réalisée en se basant sur les 8 acides aminés du site actif du domaine. Comme pour le logiciel précédent, la séquence protéique de chaque domaine peut être récupérée spécifiquement.

### **NRPS Predictor 2 : outil pour la prédiction des monomères activés par les domaines d'adénylation**

La prédiction de la composition des peptides issues de la synthèse non-ribosomique est réalisée au moyen du logiciel NRPS Predictor 2 à partir de la séquence protéique de la synthétase étudiée ou de la composition de la signature d'un domaine A à analyser. Le logiciel fournit pour chaque domaine d'adénylation une prédiction du monomère activé au moyen de différentes méthodes de prédictions avec les différents niveaux de précision « three cluster », « large cluster », « small cluster » et « single AA » (voir paragraphe 1.1.6).

### **AntiSMASH**

L'analyse de séquences génomiques au moyen du logiciel antiSMASH est réalisée en utilisant un génome entier au format EMBL ou au moyen de fragments de génomes en cours de séquençage et d'assemblage (drafts). Les séquences sont rentrées et l'analyse est réalisée pour tous les types de clusters de gènes impliqués dans la production de métabolites secondaires détectables par le logiciel, parmi lesquels les NRPS les hybrides NRPS-PKS peuvent être sélectionnés.



**Figure 35 :** Exemple de résultat d’analyse d’un génome fourni par le logiciel antiSMASH. Les résultats présentés sont de haut en bas : Une vue globale du nombre de clusters de métabolites secondaires sous la forme de disques de couleurs numérotés (la couleur indiquant la nature du métabolite produit). Une présentation des gènes du cluster sélectionné. Une présentation de l’organisation en domaines enzymatiques des protéines produites par le cluster sélectionné. Une comparaison de l’organisation globale du cluster avec des clusters identifiés chez des souches connus. Le panneau de droite présente une prédiction de la structure du peptide final produit réalisée au moyen de NRPSPredictor 2 et permet d’envoyer ces données vers la base de données NORINE pour comparer le peptide prédit avec la base de données.

Le logiciel fournit un résultat présentant les différents clusters identifiés et la nature des métabolites produits, une organisation des gènes du cluster et le rôle de chacun de ces gènes (synthèse, transport, régulation...), une organisation des domaines enzymatiques composant les protéines identifiées, ainsi qu’une comparaison du cluster identifié avec des clusters de métabolites secondaires présents chez d’autres souches séquencées. Le logiciel présente également une prédiction de la structure globale de la molécule finale produite et un lien permettant d’effectuer la comparaison de cette structure avec les peptides présents dans la base de données NORINE.

Dans le cadre de l’analyse de génomes complets, l’analyse peut être d’une durée importante (plusieurs heures). Une adresse de courrier électronique peut être fournie au logiciel afin d’être notifié de la fin de celle-ci. Les résultats obtenus au moyen d’antiSMASH peuvent être téléchargés partiellement ou intégralement au format XLS, EMBL ou Genbank.

## NORINE

La base de données de peptides non-ribosomiques NORINE est utilisée pour comparer les peptides prédits par bioinformatique aux peptides déjà identifiés dans la littérature. La comparaison d’un peptide prédit avec la base de données peut être réalisée au moyen de

différents outils fournis par NORINE. Ainsi, la base de données peut être consultée pour la recherche d'un peptide selon son nom, son organisme producteur, sa structure primaire (séquence en monomères), ses particularités de structure, telle une cyclisation ou la condensation avec un acide gras. La recherche peut également être effectuée sur la base d'une structure partielle pouvant être obtenue dans le cas où seule une partie des domaines A identifiés ont permis une prédiction du monomère activé lors de la recherche de potentiel de production de peptides non-ribosomiques chez un organisme. Pour cela, il faut ouvrir l'onglet « structure search » et ensuite dessiner le peptide ou fragment de peptide sous forme monomérique. Cette étape est facilitée par l'utilisation d'un éditeur développé à cet effet. En résultat Norine retournera la liste de tous les peptides qui contiennent le fragment demandé.

# **CHAPITRE 3**

## **Résultats**

# 1. Etude du mécanisme de synthèse de la kurstakine

## 1.1 Introduction

Parmi les lipopeptides découverts au sein du genre *Bacillus*, la kurstakine n'a pour le moment été que peu étudiée, notamment en ce qui concerne son mécanisme de production. Si les gènes de production de la kurstakine synthétase ont été identifiés, peu d'informations sont connues quant à la dynamique de production de la molécule, à son transport, ou à l'état physiologique des cellules produisant le lipopeptide.

Cependant, la kurstakine est produite en quantités relativement faibles en comparaison à d'autres lipopeptides de *Bacillus*, telles que la surfactine ou l'iturine, et le procédé de production requiert la croissance des bactéries sur un milieu de culture gélosé. L'extraction de la kurstakine s'effectue alors sur les cellules et les spores récupérées sur le milieu gélosé. Ce procédé est fastidieux et difficilement applicable à plus grande échelle.

Le but des travaux décrits dans cette partie est d'apporter des éléments de compréhension des facteurs limitant la production de la kurstakine par *Bacillus thuringiensis*, afin de développer un procédé de production en culture liquide, facilitant ainsi la purification et le dosage du lipopeptide. Dans ce but l'un des premiers objectifs sera d'identifier une ou plusieurs souches avec lesquelles les travaux seront réalisés. Les critères principaux de sélection seront basés sur les connaissances de ces souches au niveau génomique et leur taux de production en kurstakine.

L'effet d'un ensemble de facteur limitant potentiellement la production du lipopeptide sera également étudié par l'utilisation de différents mutants de ces souches réalisés par le groupe partenaire de ces travaux, l'équipe MICALIS de l'INRA. La production du lipopeptide par ces mutants sera analysée afin de mieux comprendre l'implication de certaines voies métaboliques de *B. thuringiensis* dans la production de kurstakine. La cinétique de production du lipopeptide sera également étudiée en milieu de culture solide et liquide afin de mieux comprendre l'évolution de cette production dans le temps.

Enfin, ces travaux seront complétés par une étude d'activités biologiques potentiellement valorisable de la kurstakine : les activités antifongiques et antibactériennes de la molécule.

## 1.2 Etude du mécanisme de production de la kurstakine

### 1.2.1 Détection de la kurstakine produite par *Bt HD-73* et *Bt407*

Dans le cadre de ces travaux, les souches *B. thuringiensis* HD-73 et Bt407 ont été étudiées pour la production de kurstakine. Ces deux souches présentent plusieurs intérêts dans le cadre de cette étude : plusieurs mutants de la souche *B. thuringiensis* Bt407 avaient auparavant été construits par l'équipe MICALIS de l'INRA de Jouy-en-Josas et étaient disponibles au début de ces travaux, tandis que la souche HD-73, également disponible, fournissait un élément de comparaison et de confirmation des résultats de production de kurstakine obtenus avec les modifications réalisées sur la souche Bt407.

Afin de confirmer le potentiel de production de kurstakine par les souches Bt HD-73 et Bt407, une analyse en spectrométrie de masse des cellules est réalisée pour détecter le lipopeptide. En effet, la présence de gènes de synthétase ne signifie pas pour autant que la production du peptide ou lipopeptide associé sera effective. Ainsi, plusieurs souches de *B. thuringiensis* chez lesquelles les gènes de la kurstakine synthétase avaient été détectés ne présentent pas de production du lipopeptide lors d'une analyse par spectrométrie de masse. (Abderrahmani *et al.*, 2011).

La production de kurstakine par la souche *B. thuringiensis* HD-73 a été analysée par spectrométrie de masse MALDI-TOF sur colonies cultivées sur milieu LB pendant 24 heures (Figure 36).

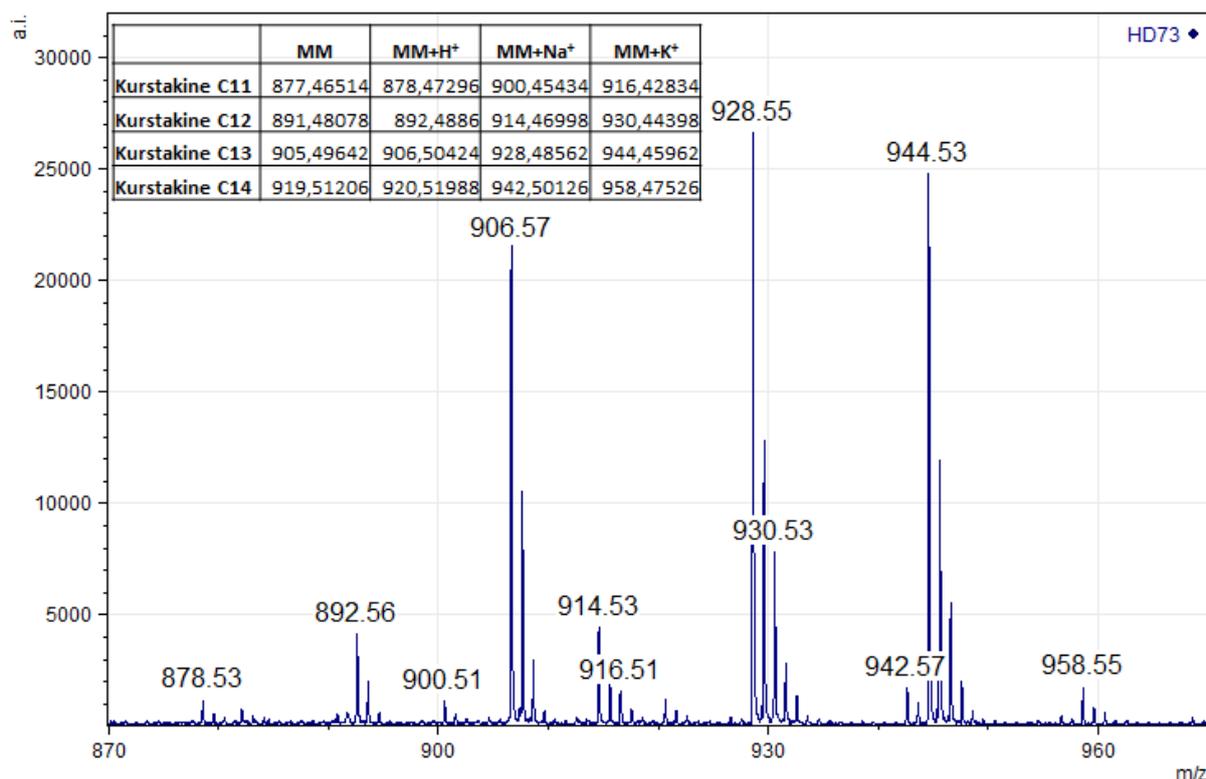


Figure 36 : Analyse par spectrométrie de masse MALDI-TOF d'une colonie de *B. thuringiensis* HD-73 cultivée sur milieu LB gélosé pendant 24 heures à 30°C. Une colonie a été reprise dans 20µL de matrice HCCA, le mélange ayant été ensuite brièvement centrifugé. L'analyse est réalisée sur le surnageant contenant les molécules récupérées de la surface des cellules. Le tableau rappelle les valeurs m/z des adduits H<sup>+</sup>, Na<sup>+</sup> et K<sup>+</sup> des différentes formes observées.

Les différents pics détectés correspondent aux adduits H<sup>+</sup>, Na<sup>+</sup> et K<sup>+</sup> de la kurstakine portant un acide gras à 11, 12, 13 et 14 carbones, indiquant la production de ces 4 formes de lipopeptide par la souche *B. thuringiensis* HD-73.

L'intensité relative des groupes de pics semblent indiquer la production d'une forte majorité de kurstakine C13.

De la même manière, une analyse des formes de kurstakines produites par la souche *B. thuringiensis* Bt407 est réalisée par spectrométrie de masse dans les mêmes conditions (Figure 37).

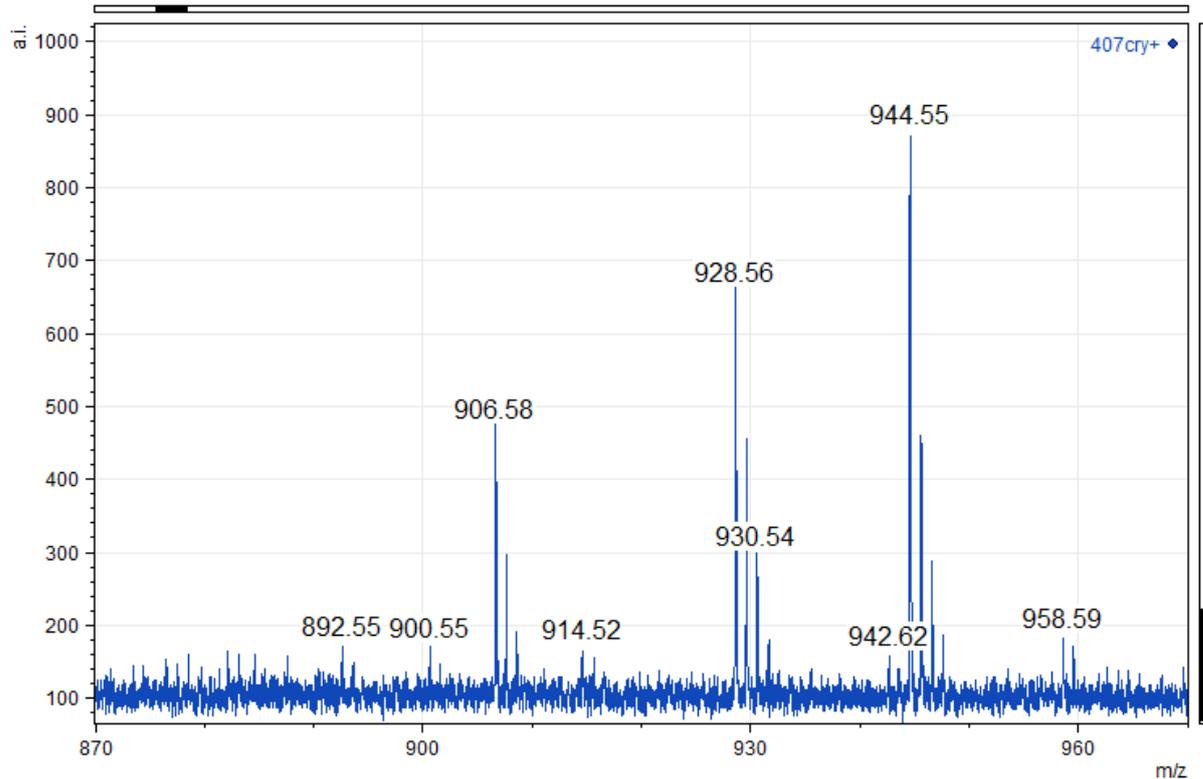


Figure 37 : Analyse par spectrométrie de masse MALDI-TOF d'une colonie de *B. thuringiensis* Bt407 cultivée sur milieu LB gélosé pendant 24 heures à 30°C. Une colonie a été reprise dans 20µL de matrice HCCA, le mélange ayant été ensuite brièvement centrifugé. L'analyse est réalisée sur le surnageant contenant les molécules récupérées de la surface des cellules. Le tableau rappelle les valeurs m/z des adduits H<sup>+</sup>, Na<sup>+</sup> et K<sup>+</sup> des différentes formes observées.

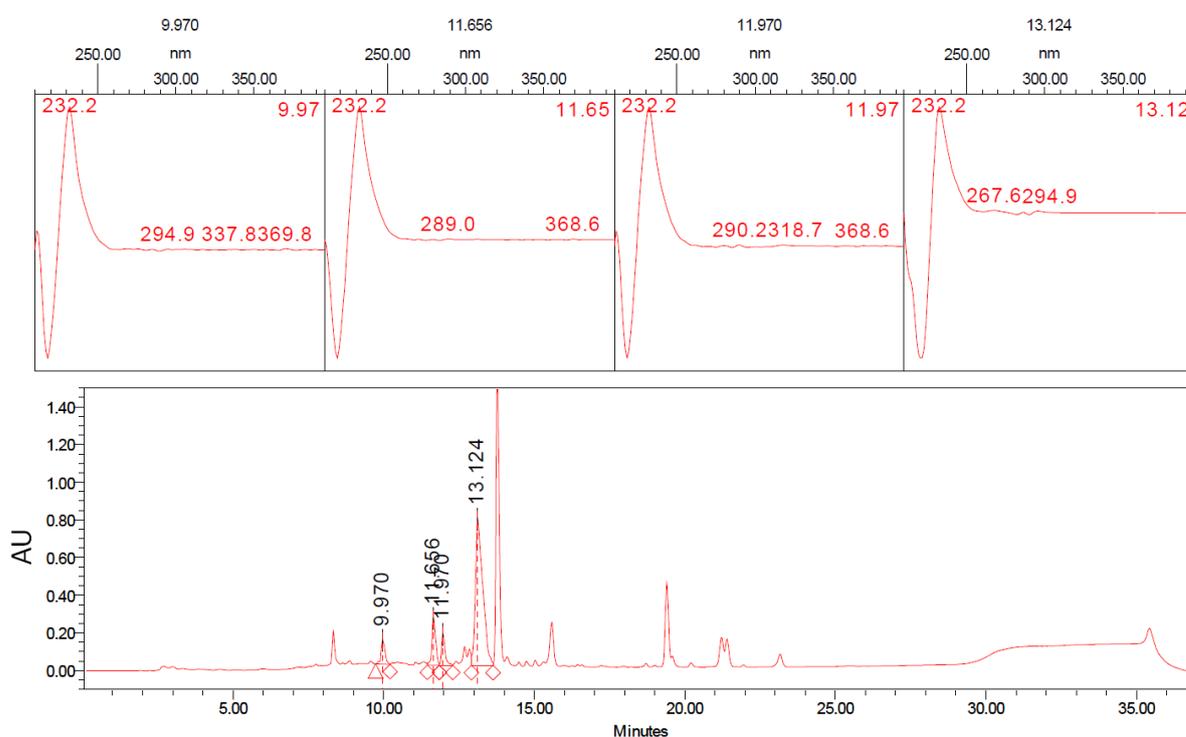
La souche *B. thuringiensis* Bt407 présente un profil de pics proche de celui de la souche *B. thuringiensis* HD-73. Les adduits H<sup>+</sup>, Na<sup>+</sup> et K<sup>+</sup> de la forme C13 sont majoritaires, et les pics des adduits des formes C12 et C14 ne sont que peu visibles. En revanche, seul l'adduit H<sup>+</sup> de la forme C11 peut être détecté. L'intensité des pics observés étant moins forte que celle obtenue avec la souche *B. thuringiensis* HD-73, le signal des pics correspondant aux autres adduits n'est sans doute pas suffisant pour se dégager du bruit de fond.

Les deux souches analysées semblent donc produire un mélange similaire des formes C11 à C14 de la kurstakine.

### 1.2.2 Analyse quantitative de production de kurstakine par différentes souches de différentes souches de *B. thuringiensis* et *B. cereus*

Les gènes de la kurstakine synthétase et la production de kurstakine ayant été observées chez différentes souches, l'identification de la souche présentant le meilleur taux de production en lipopeptide est nécessaire pour la sélection d'une souche d'intérêt biotechnologique. Afin de déterminer ce potentiel, les taux de production de kurstakine de

différentes souches sauvages de *B. thuringiensis* et ont été évalués en réalisant le dosage HPLC d'extraits préparés à partir de cultures réalisées en milieu solide AK agar selon le protocole modifié d'Hathout *et al.* décrit dans la section matériels et méthodes. Les souches sont cultivées sur le milieu AK agar une nuit à 37°C et 3 jours à températures ambiantes. Après extraction et purification des lipopeptides sur colonne SPE C8, les échantillons sont analysés par HPLC permettant la séparation des différentes formes de la kurstakine par une colonne greffée par de la phase C8 lors d'un gradient de solvant de 30 à 70% d'acétonitrile. La séparation des molécules est suivie par la mesure de l'absorbance à 214 nm. Un exemple de chromatogramme obtenu est présenté ci-dessous (Figure 38).



**Figure 38 : Chromatogramme HPLC de 20 $\mu$ L d'un extrait liquide enrichi en kurstakine produite par Bt HD-73 en milieu AK agar après 1 nuit à 37°C et 3 jours à température ambiante selon le protocole décrit par Hathout *et al* (2000). La kurstakine présente 4 pics à 9,97, 11,65, 11,97 et 13,12 minutes d'analyse. La séparation est effectuée sur une colonne Kinetex C8, avec un gradient de concentration en acétonitrile de 30 à 70% de 0 à 25 minutes.**

L'analyse HPLC permet d'identifier 4 pics aux temps de rétentions 9,97, 11,65, 11,97 et 13,12 minutes. Les trois premiers pics présentent des intensités du même ordre de grandeur. Le pic à 13,12 minutes présente une intensité relativement plus importante.

Cette analyse est comparée avec les données de la littérature. La première analyse d'extraits de spores de Bt par HPLC a été réalisée en 2000 par Hathout *et al.*. Elle a démontré que dans les conditions utilisées, la kurstakine est détectée par HPLC sous la forme de 4 pics entre 9 et 14 minutes d'élution (Figure 39).

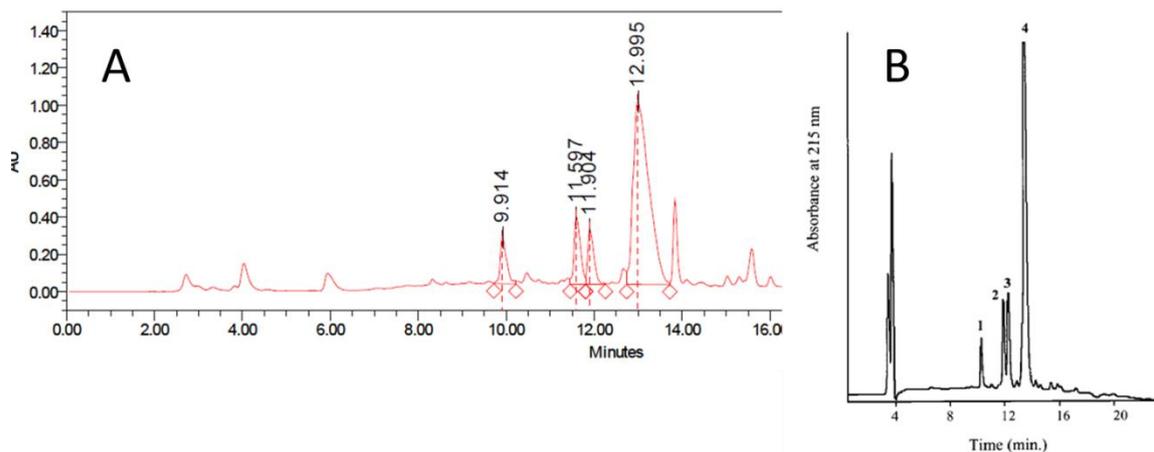


Figure 39 : Comparaison du profil de chromatogramme HPLC obtenu lors de cette étude pour production de kurstakine par la souche *B. thuringiensis* HD-1 sur milieu AK agar(A) avec l'analyse réalisée lors de la découverte de la kurstakine (B) (Hathout *et al.*, 2000).

La comparaison des résultats obtenus avec le premier chromatogramme de séparation des différentes formes de kurstakine avec les résultats obtenus lors de cette étude permet d'observer la présence du même massif de 4 pics entre 9 et 14 minutes d'éluion.

De plus, une confirmation de la nature des pics observés est réalisée en récoltant le produit de l'éluion à la sortie du système HPLC entre 9 et 14 minutes. Après concentration de l'extrait par évaporation du solvant, le produit séché est repris dans un volume de 20  $\mu$ l de matrice HCCA pour une analyse en spectrométrie de masse MALDI-TOF (Figure 40).

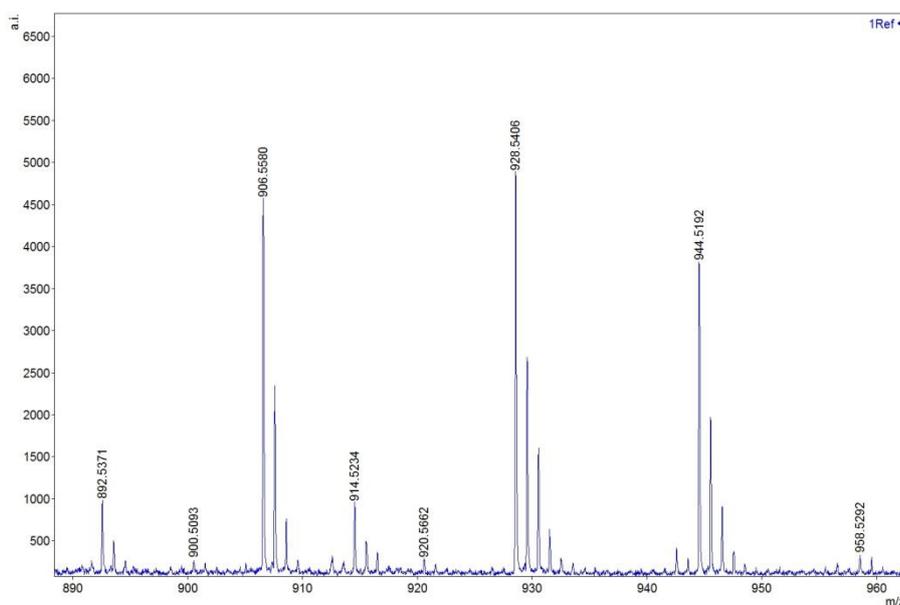


Figure 40 : Analyse par spectrométrie de masse MALDI-TOF des pics aux temps de rétention compris entre 9 et 14 minutes lors de l'analyse HPLC de l'extrait de kurstakine produit par *B. thuringiensis* HD-73 sur milieu AK agar.

L'analyse par spectrométrie de masse MALDI-TOF des pics récoltés lors de l'analyse HPLC permet d'identifier les différents adduits  $H^+$ ,  $Na^+$  et  $K^+$  les formes C11 à C14 de la kurstakine, confirmant ainsi que les pics détectés par HPLC correspondent aux différentes formes du lipopeptide.

L'utilisation de la technique HPLC permettant de détecter les différentes formes de kurstakines, le dosage de la production du lipopeptides par les souches *B. thuringiensis* HD-1, HD-73 et Bt407 a été réalisé selon le même protocole (Figure 41). Aucun standard de kurstakine n'étant disponible à ce jour, un standard de surfactine, possédant le même nombre de liaisons peptidiques que la kurstakine, a été utilisé pour réaliser cette quantification.

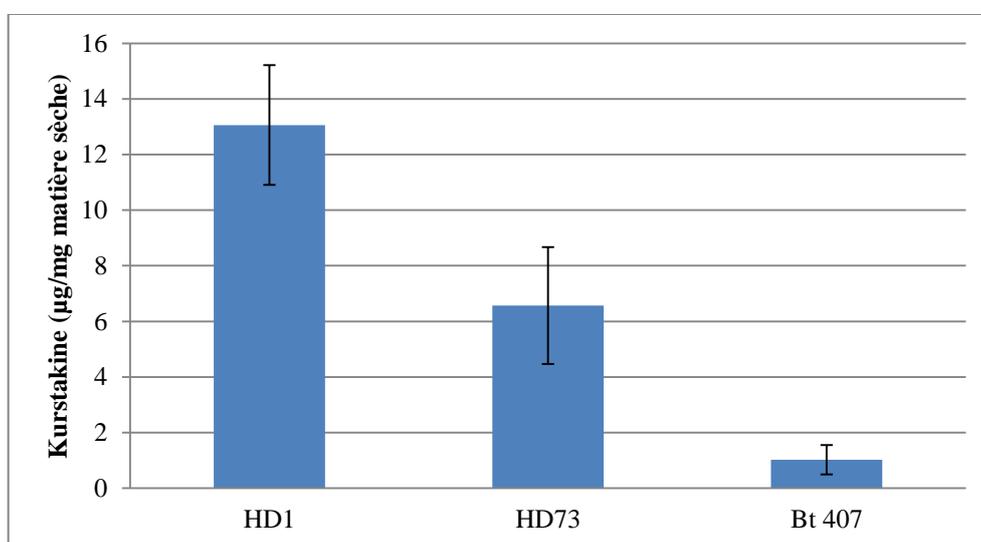


Figure 41 : Taux de production de kurstakine sur milieu AK agar après 1 nuit à 37°C et 3 jours à température ambiante par les souches de *Bacillus thuringiensis* HD-1, HD-73 et Bt407.

La souche *B. thuringiensis* HD-73 produit environ 6 µg de kurstakine par mg de matière sèche. A titre de comparaison la souche *B. thuringiensis* HD-1 produit entre 15 et 20 µg/mg de matière sèche. La souche *B. thuringiensis* Bt407 produit quant à elle 1µg/mg de matière sèche.

A la vue de ces résultats la souche *B. thuringiensis* HD-1 présente le taux de production de kurstakine le plus élevé parmi les souches testées, suivi de *B. thuringiensis* HD-73 et *B. thuringiensis* Bt407. Néanmoins, lors du début des travaux réalisés dans le cadre de cette thèse, seuls les génomes des souches *B. thuringiensis* HD-73 et Bt407 avaient été séquencés et publiés. La souche HD-73 possédant le meilleur taux de production du lipopeptide entre ces

deux souches, elle a été majoritairement utilisée pour l'étude des facteurs limitant potentiellement la production de kurstakine.

Un ensemble de mutants réalisés à partir de la souche *B. thuringiensis* 407 étant également disponibles au début de ce projet, ceux-ci ont été également utilisés pour l'analyse de l'impact de certains de ces facteurs.

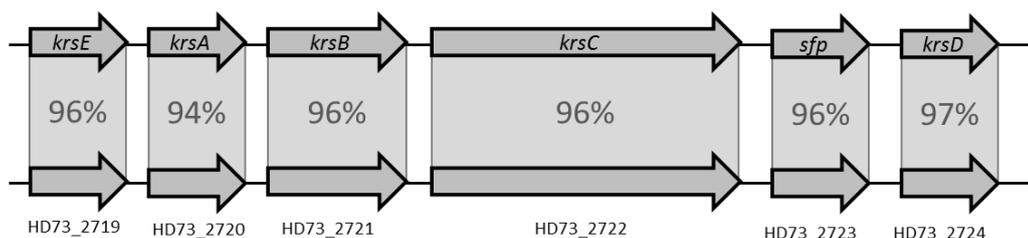
### *1.2.3 Analyse bioinformatique de la kurstakine synthétase chez Bacillus thuringiensis HD-73*

#### *1.2.3.a Identification des gènes impliqués dans la synthèse de la kurstakine*

Afin de valider l'intérêt de l'utilisation de la souche *B. thuringiensis* HD-73 pour l'étude du mécanisme de production de la kurstakine, le génome de la souche a été analysé au moyen des outils dédiés à l'analyse des NRPS décrits dans la partie Matériels et Méthodes, pour identifier les gènes responsables de la production de la kurstakine chez cette souche. L'analyse du génome de *B. thuringiensis* HD-73 par le logiciel AntiSMASH a révélé la présence de plusieurs clusters de gènes potentiellement impliqués dans la production de métabolites secondaires. Parmi ceux-ci, 4 clusters de gènes codant pour des NRPS ont été détectés. Parmi ces clusters, un seul (le cluster 5) code pour une synthétase contenant le nombre de domaines d'adénylation nécessaires pour la synthèse de la kurstakine. Celui-ci est composé de 3 gènes principaux NRPS : HD73\_2720, contenant 1 module (C-A-T-E), HD73\_2721 contenant 2 modules (C-A-T-C-A-T), et HD73\_2722 contenant 4 modules (C-A-T-C-A-T-C-A-T-E-C-A-T-Te). Un gène codant pour une phosphopantéthéinyl transférase (HD73\_2723) est également détecté en aval de ces trois gènes lui-même suivi par un gène codant pour une deuxième thioestérase (HD73\_2724). En amont de ces cinq gènes, un gène de transport (HD73\_2719), codant pour une protéine de la famille des MFS-1 (Major Facilitator Superfamily) a été observé.

Les gènes du cluster 5 identifiés chez Bt HD-73 ont été alignés aux gènes responsables de la production de la kurstakine synthétase chez *B. thuringiensis* serovar pondicheriensis BGSC4BA1. L'homologie entre les gènes chez les deux organismes varie de 94 à 97%.

*Bacillus thuringiensis* serovar pondicheriensis BSCG 4BA1



*Bacillus thuringiensis* HD-73

Figure 42 : Pourcentage d'identité entre les gènes de la kurstakine synthétase chez *B. thuringiensis* serovar pondicheriensis BSCG 4BA1 et les gènes du cluster 5 détecté chez *B. thuringiensis* HD-73

Le fort pourcentage d'identité des différents gènes trouvés dans le cluster 5 chez Bt HD-73 avec les gènes de la kurstakine synthétase de *B.thuringiensis* serovar pondicheriensis, ainsi que la similarité de l'organisation des gènes et des domaines NRPS entre les deux souches indique que les deux synthétases sont proches. La synthétase détectée chez *B. thuringiensis* HD-73 est donc probablement impliquée dans la production de kurstakine.

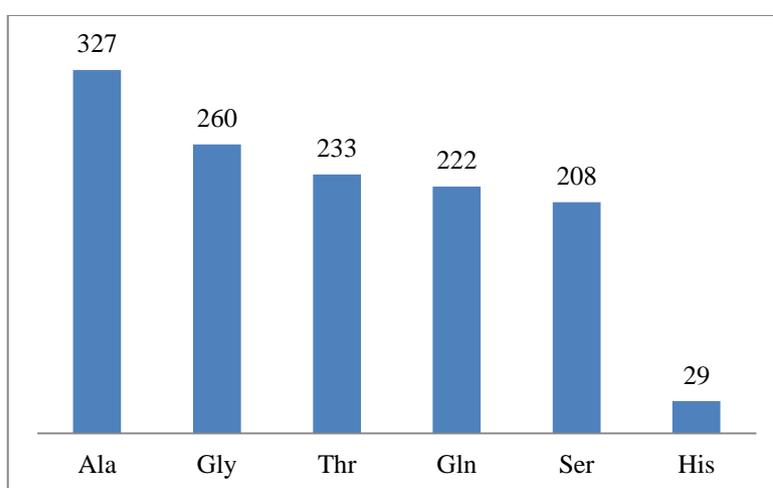
1.2.3.b Prédiction du produit de la synthétase identifiée chez *B. thuringiensis* HD-73

Afin de confirmer la nature du produit final de la synthétase identifiée chez la souche Bt HD-73, la prédiction des monomères activés par les domaines A de cette synthétase est réalisée au moyen du logiciel NRPS Predictor 2 du modèle Nearest Neighbor, permettant alors d'obtenir pour chaque domaine A la prédiction du monomère activé. La qualité de la prédiction est évaluée au moyen du score associé à chaque prédiction.

Tableau 5 : Prédiction des monomères activés par les domaines A de la kurstakine synthétase chez *B. thuringiensis* HD-73 et *B thuringiensis* serovar pondicheriensis réalisés au moyen de NRPS Predictor 2 avec le modèle Nearest Neighbour. Le score associé à chaque prédiction est indiqué entre parenthèses. La séquence en acides aminés de la kurstakine est rappelée en dernière ligne.

Domaines	A1	A2	A3	A4	A5	A6	A7
<b>Bt HD-73</b>	Thr (100%)	Gln (70%)	Ala (100%)	Ser (100%)	Pro (70%)	Gln (100%)	Gln (90%)
<b>Bt pondicheriensis</b>	Thr (100%)	Gln (70%)	Ala (100%)	Ser (100%)	Pro (70%)	Gln (100%)	Gln (90%)
<b>Kurstakine</b>	Thr	Gly	Ala	Ser	His	Gln	Gln

Les monomères prédits pour chaque domaine d'adénylation identifié dans la kurstakine synthétase de *B. thuringiensis* HD-73 sont Thr-Gln-Ala-Ser-Pro-Gln-Gln. Seuls 2 monomères prédits ne correspondent pas à la séquence peptidique connue de la kurstakine, à savoir la glutamine en position 2 et une proline prédite en position 5. Néanmoins, les scores associés à ces prédictions sont relativement faibles avec le modèle Nearest Neighbor (70%, contre 90% et 100% pour les monomères prédits aux autres positions). Une évaluation de la représentation de chaque monomère composant la kurstakine dans la composition des peptides référencés à ce jour a été effectuée au moyen des données de la base de données NORINE.



**Figure 43 : Comparaison du nombre de peptides référencés contenant chaque monomère inclus dans la séquence peptidique de la kurstakine. (Source : NORINE)**

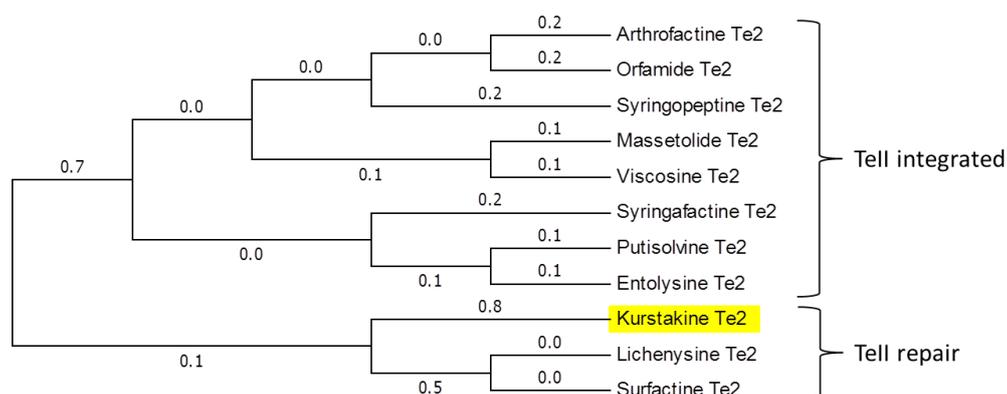
Parmi les 6 monomères composant la kurstakine, 5 sont retrouvés dans plus de 200 peptides, tandis que l'histidine n'est retrouvée que chez 29 peptides référencés dans la base de données NORINE. Ce faible nombre de peptides incluant l'histidine dans leur composition entraîne la connaissance d'un faible nombre de domaines d'adénylation activant ce monomère, et donc une mauvaise connaissance de la signature correspondante. Ainsi le faible score associé à la prédiction en position 5 peut être expliqué par la faible représentation de l'histidine parmi les monomères rentrant dans la composition des peptides non-ribosomiques.

La très forte identité entre les gènes NRPS trouvés chez *B. thuringiensis* HD-73 et ceux connus de la kurstakine synthétase chez *B. thuringiensis* serovar pondicheriensis BGSC 4BA1 ainsi que la similitude des prédictions des monomères incorporés par chaque domaine permettent d'identifier le cluster de gènes NRPS identifié chez Bt HD-73 comme très potentiellement codant de la kurstakine synthétase.

Parmi les autres clusters NRPS détectés, aucun ne correspond à un groupe de gènes connus pour la production de lipopeptide, semblant indiquer que la kurstakine est la seule famille de lipopeptides produite par la souche Bt HD-73.

### 1.2.3.c Fonction de la deuxième thioestérase

Comme indiqué dans la synthèse bibliographique, une deuxième thioestérase peut parfois jouer un rôle important dans l'activité d'une NRPS. Le dernier gène du cluster identifié, présentant 97% d'identité avec le gène *krsD* de BSCG 4BA1, code comme ce dernier une deuxième thioestérase (TeII), la première étant codée par l'extrémité 3' du gène *krsC*. L'étude phylogénétique de la séquence du gène de cette thioestérase (Figure 44) permet d'attribuer un rôle potentiel à cette TeII, notamment dans le cadre de la production de kurstakine. Pour cela, les séquences de gènes codant pour des TeII connues ont été collectées. Des TeII possédant les deux types d'activité possible (TeII integrated, dont le rôle est proche des thioestérases de type I, permettant la cyclisation potentielle et le relargage du peptide produit, et TeII repair impliquée dans la réparation des domaines T ayant chargé un mauvais co-facteur (voir paragraphe 1.1.4.5) ont été collectées à partir de la base de données NCBI. Les séquences ont ensuite été alignées au moyen du logiciel MUSCLE. L'alignement a ensuite été utilisé pour la création d'un arbre phylogénétique permettant d'attribuer à la TeII de la kurstakine une fonction potentielle.



**Figure 44 : Arbre phylogénétique de l'alignement de domaines TeI, TeII integrated et TeII repair de NRPS connues. La thioestérase II de la kurstakine synthétase a été ajoutée à l'alignement.**

A la vue de l'arbre phylogénétique obtenu, la TeII issue du gène *krsD* est proche des thioestérase de type II repair.

Si les TeII « repair » ne sont pas strictement indispensables au fonctionnement des NRPS, elles jouent tout de même un rôle important dans l'efficacité de l'enzyme. En effet, le chargement d'un mauvais co-facteur par les domaines de thiolation signifie que l'enzyme, sous sa forme apo-enzyme est non fonctionnelle, et requiert l'action de la TeII pour passer sous sa forme holo-enzyme active. L'interruption du gène codant pour la TeII repair de la surfactine synthétase mène ainsi à une baisse de 84% de la production du lipopeptide (Koglin *et al.*, 2008 ; Schneider et Marahiel., 1998), tandis qu'une délétion de la TeII repair de l'arthrofactine entraîne une baisse de la production de 95% (Roonsgawang *et al.*, 2007).

Le rôle de la deuxième thioestérase, codée par *krsD*, est donc important pour le maintien d'un taux de production important en lipopeptide. Ainsi, dans le cadre de l'analyse de génomes bactériens pour la recherche d'organismes potentiellement producteur de kurstakine, la présence et l'intégrité du gène *krsD* dans le génome de souche étudié doit être examiné avec attention.

#### 1.2.4 Cinétique de production de la kurstakine

Lors de la découverte de la kurstakine en 2000, ce lipopeptide avait été identifié à partir de spores de *B. thuringiensis*, obtenues après 4 jours de culture sur milieu de culture AK agar#2, milieu spécifiquement dédié à la production de spores bactériennes (Hathout *et al.*, 2000).

Afin de mieux comprendre les raisons qui ont amené la découverte de la kurstakine à des temps de culture très avancés, la cinétique de production de cette molécule a été suivie dans le cadre de production en milieu liquide dans le cadre de cultures en fioles erlenmeyer et en fermenteur, et en milieu solide.

Dans le cadre de l'étude de la production en milieu liquide, la souche BT HD-73 est cultivée en milieu LB et « AK » liquide pendant 72 heures à 30°C sous une agitation à 160 rpm. Le suivi de croissance est réalisé par mesure de l'absorbance à 600nm (Figure 45), et la production de kurstakine est réalisée sur chaque échantillon selon le protocole décrit dans la partie matériels et méthodes.

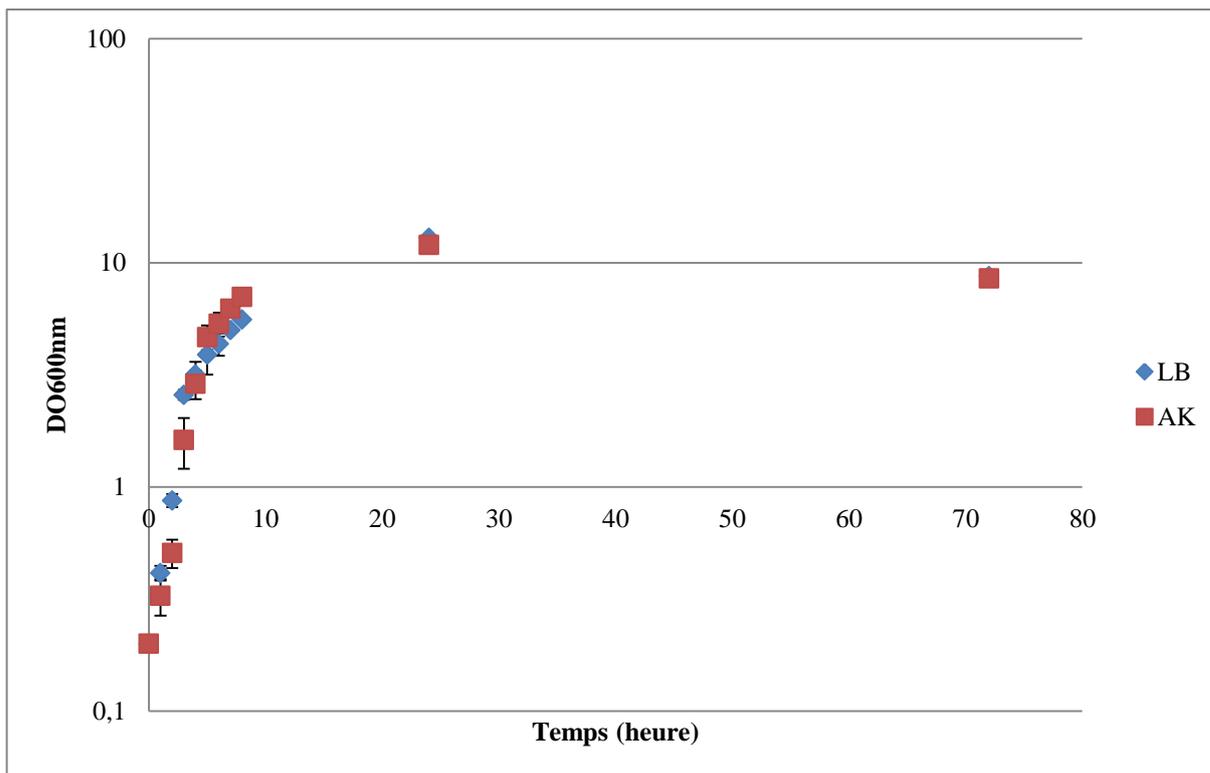
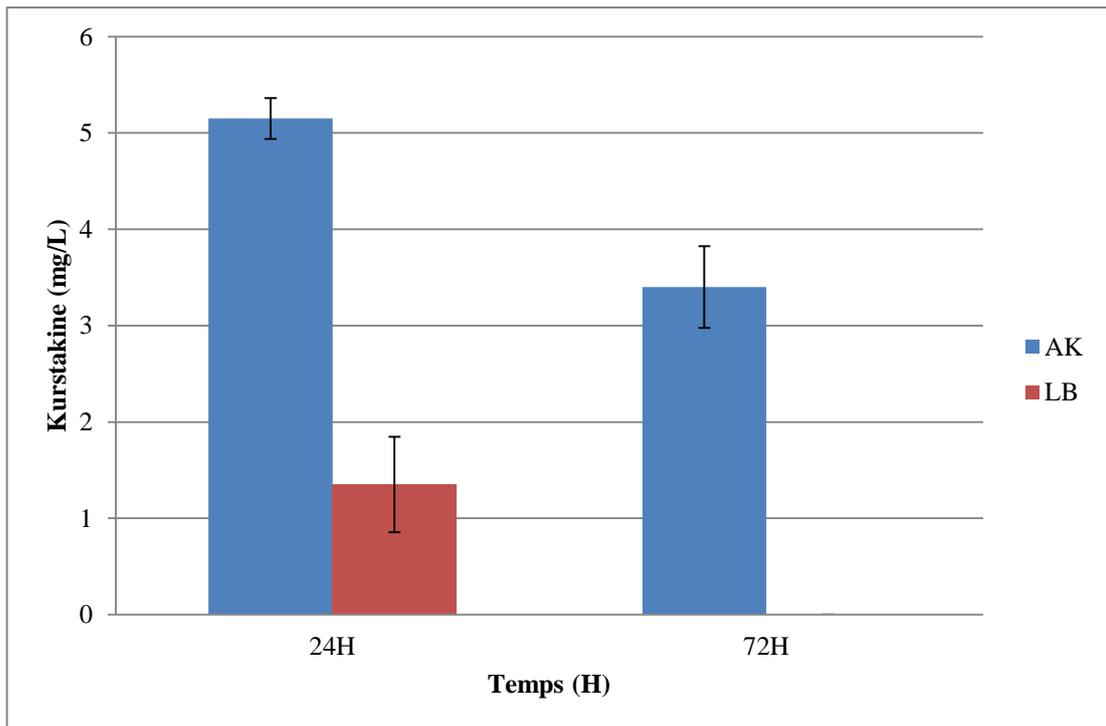


Figure 45 : Croissance de la souche Bt HD-73 en milieu liquide LB et « AK » à 30°C, 160 rpm.

La croissance de la souche Bt HD-73 est similaire dans les deux milieux de culture liquides testés. La souche connaît une phase exponentielle de croissance de 3 heures, puis connaît une phase de ralentissement de 3 à 24 heures de culture, moment à laquelle elle entre en phase stationnaire.

La cinétique de production en milieu liquide ne permet pas de quantifier de kurstakines par HPLC, aussi bien au niveau des culots cellulaires qu'au niveau des surnageants de cultures pendant les 8 premières heures de culture. En revanche, une quantification de kurstakine a pu être réalisée de façon plus tardive dans les surnageants de culture après 24h et 48h de culture (Figure 46).



**Figure 46 : Concentration en kurstakine mesurée dans le surnageant de culture de la souche Bt HD-73 cultivée en milieu LB et AK à 24 et 72H.**

Après 24 h de culture, la production de kurstakine en milieu « AK » atteint 5 mg/L dans le surnageant de culture. Après 72 h de culture, 3,7 mg/L sont mesurés. En milieu LB, 1 mg/L est mesuré après 24 h, mais la kurstakine n'est plus détectable après 72 h.

Quel que soit le milieu de culture utilisé, la production de kurstakine semble suivre le même comportement, avec une impossibilité de détecter la molécule dans le surnageant pendant les premières heures de culture, une production détectable de quelques milligrammes par litre après 24h de culture, puis une diminution de la quantité de lipopeptides mesurée dans le surnageant après 72h.

Dans le cadre des travaux réalisés, une première étude de la production de kurstakine par *B. thuringiensis* en fermenteur a été menée. La souche *B. thuringiensis* HD-73 a été cultivée dans 3L milieu LB pendant 24H en batch. La croissance de la culture a été suivie par mesure de l'absorbance à 600nm (Figure 47). Toutes les heures, un prélèvement de 20mL a été effectué, et le culot et le surnageant ont été analysés séparément après leur séparation par centrifugation.

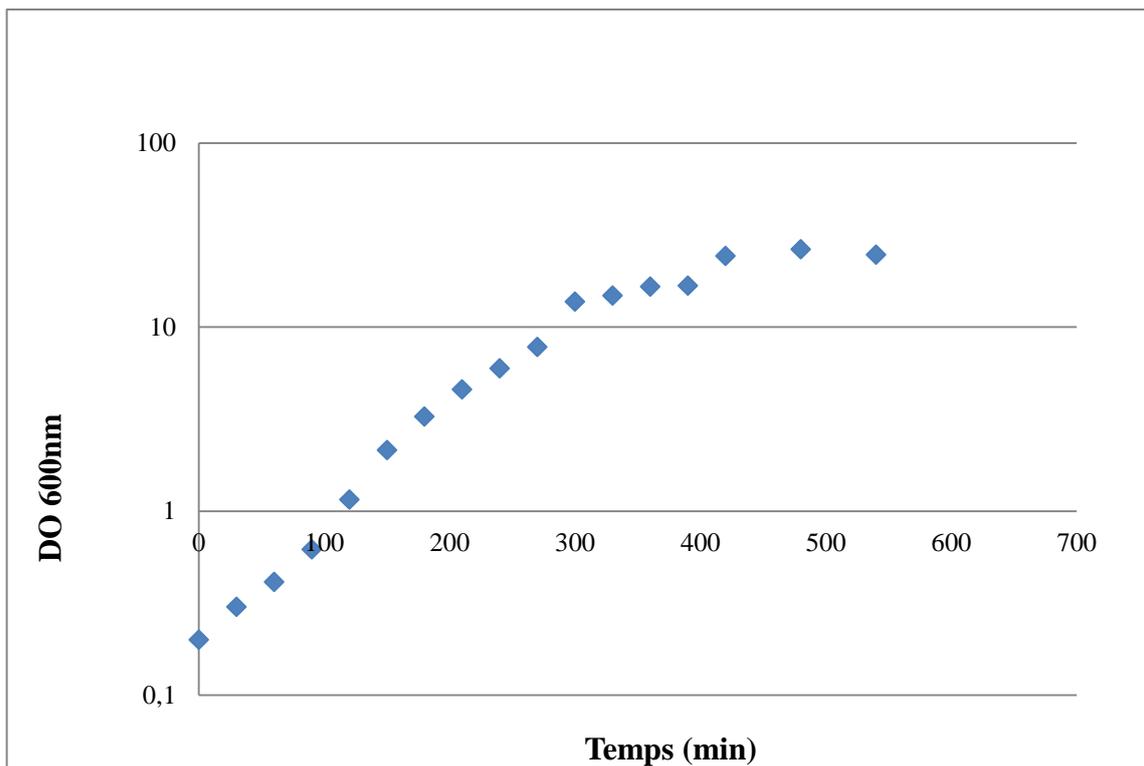


Figure 47 : Courbe de croissance de la souche *Bacillus thuringiensis* HD-73 en fermenteur de 3L en milieu LB

*Bacillus thuringiensis* présente en fermenteur une phase de croissance exponentielle de 180 minutes, et une phase de ralentissement de 180 à 420 minutes, après laquelle la souche entre en phase stationnaire.

Le prélèvement réalisé 8h après le début de la culture permet d'identifier par HPLC un pic correspondant à la kurstakine, avec un temps de rétention de 13 minutes environ, et un profil d'absorbance en UV similaire à celui observé lors des quantifications du lipopeptide réalisées précédemment (voir paragraphe 1.2.2).

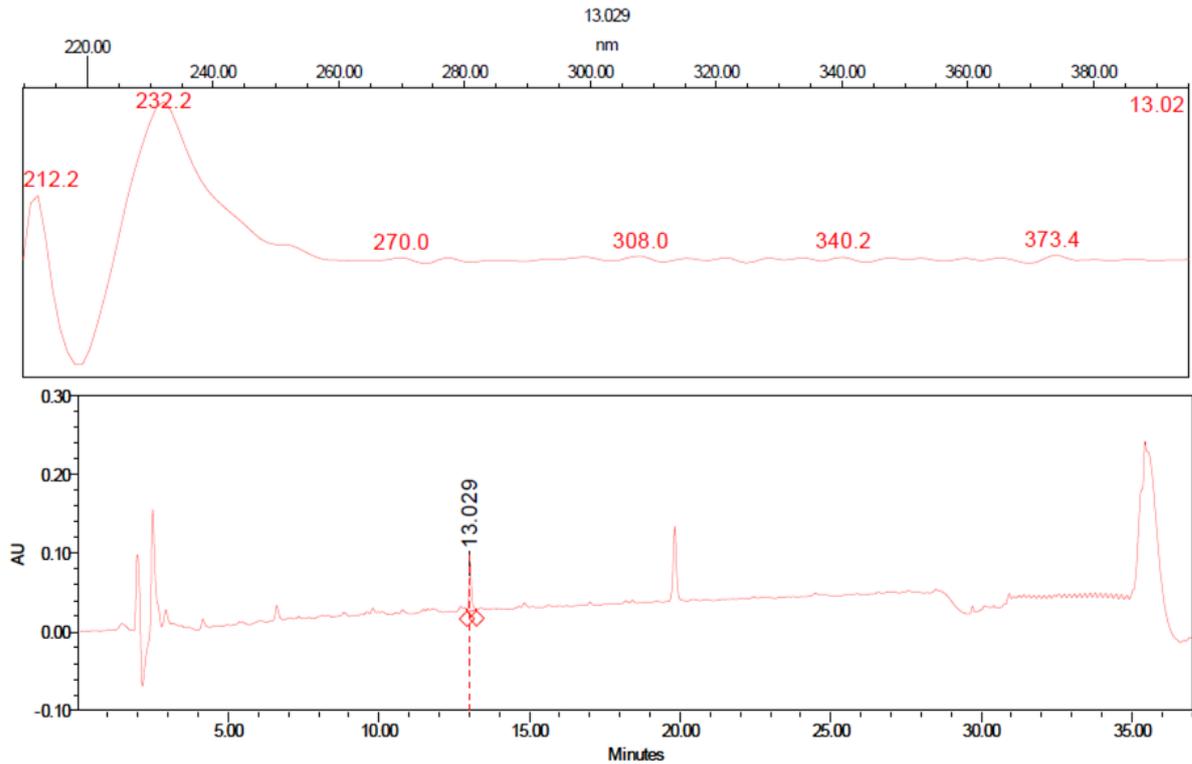
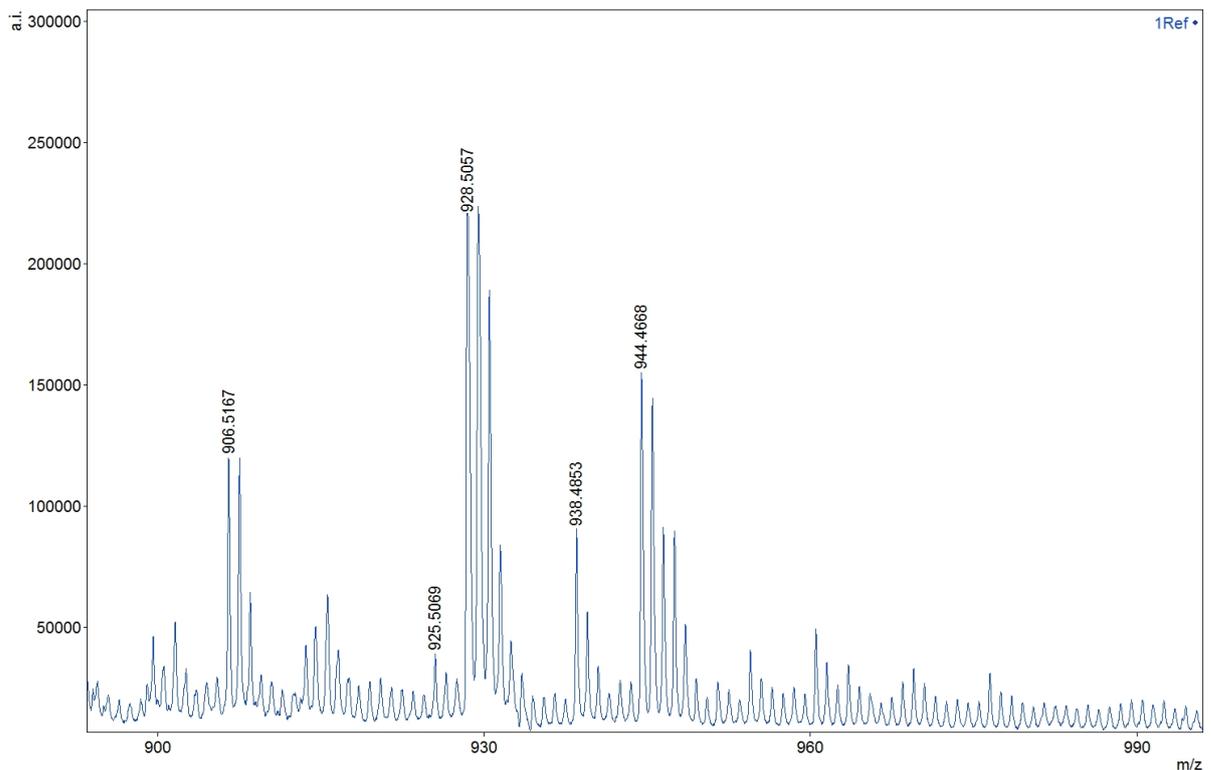


Figure 48 : Chromatogramme HPLC de l'extrait réalisé à partir du culot cellulaire de *Bacillus thuringiensis* HD-73 après 8H de culture en fermenteur en milieu LB à 30°C.

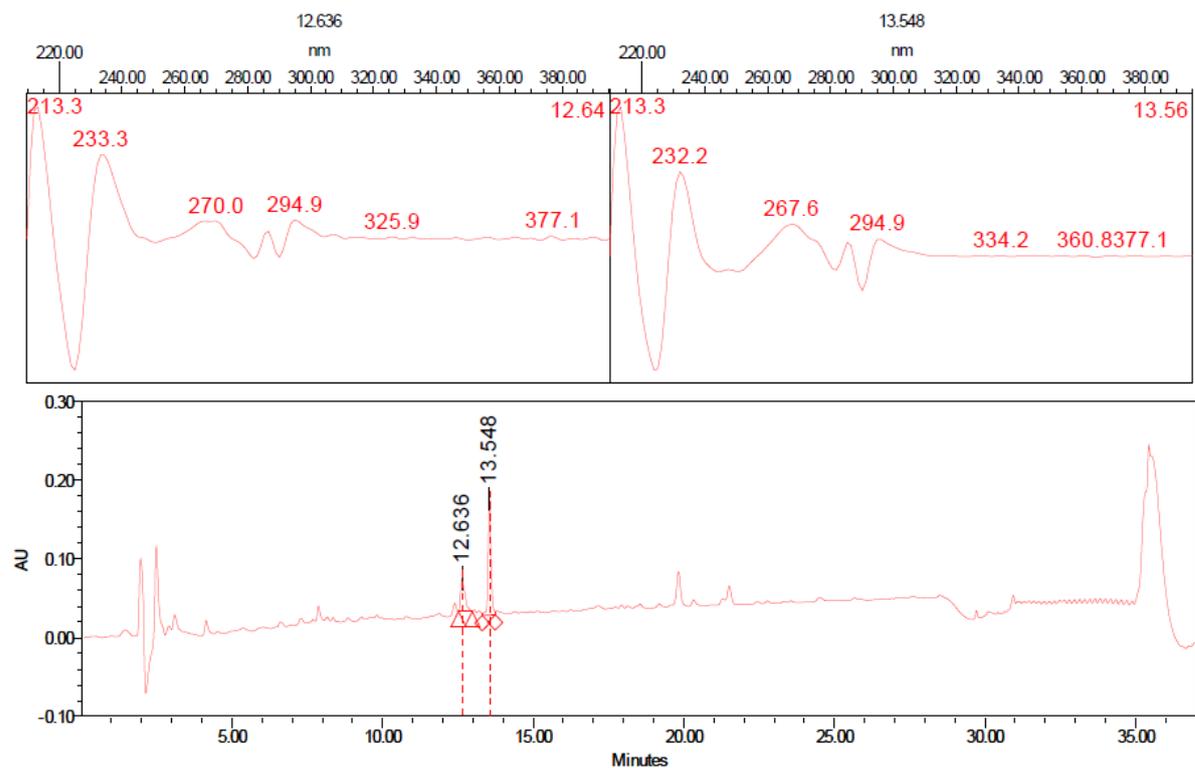
Afin de confirmer la nature des peptides élués lors de cette chromatographie, une récolte du pic a été effectuée, et l'échantillon a été concentré par réduction du volume à 10µL final par séchage, puis additionné de 10µL de matrice HCCA pour analyse par spectrométrie de masse MALDI-TOF (Figure 49).



**Figure 49 : Analyse par spectrométrie de masse MALDI-TOF du composé récolté après séparation par HPLC provenant de l'extrait cellulaire réalisé sur un culot de *B. thuringiensis* HD-73 après 8H de culture en milieu LB à 30°C.**

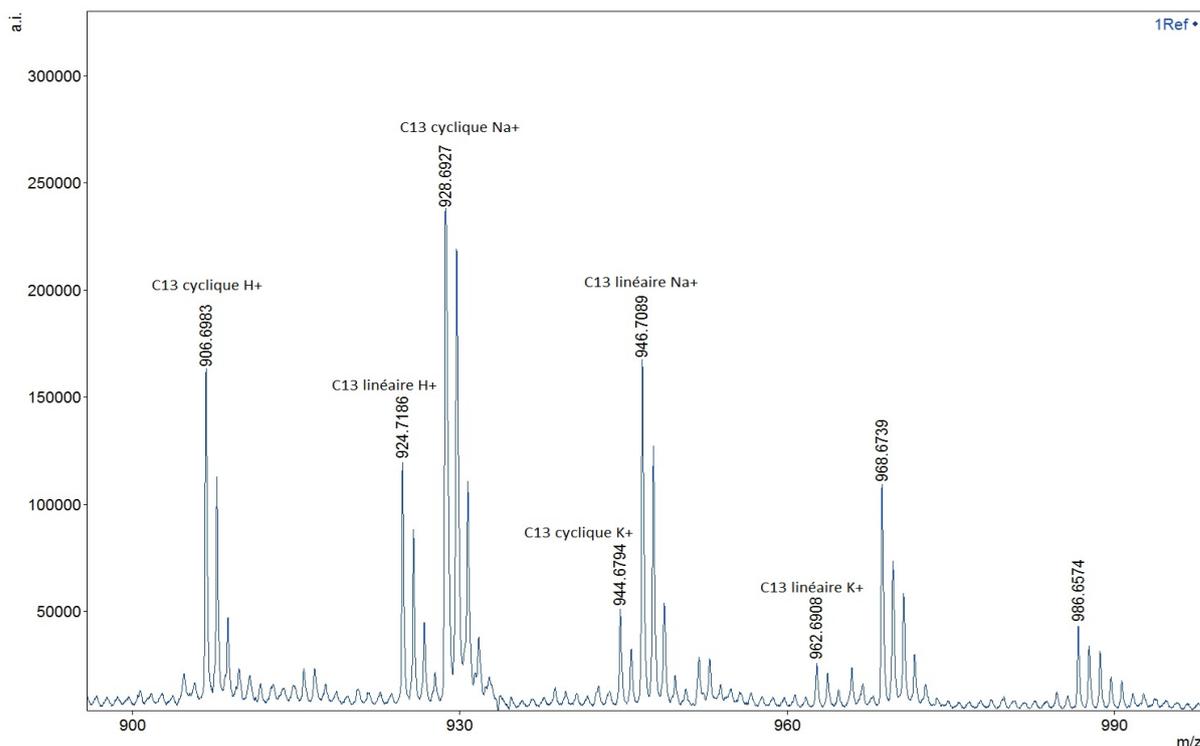
L'analyse par spectrométrie de masse permet d'observer la présence d'adduits de kurstakines de formes C13 (Figure 49).

Après 1h de culture supplémentaire, le profil HPLC observé est légèrement différent. Deux pics sont observés en HPLC, à respectivement 12,6 et 13,6 minutes. Le profil de l'absorbance en UV est différent de celui du pic observé après 8h de culture (Figure 50).



**Figure 50 :** Chromatogramme HPLC de l'extrait réalisé à partir du culot cellulaire de *Bacillus thuringiensis* HD-73 après 9H de culture en fermenteur en milieu LB à 30°C.

Une collecte a été réalisée à l'issue de la séparation par HPLC de 12 à 14 minutes d'éluion afin d'identifier les composés détectés. Le volume de l'échantillon prélevé est réduit à un volume de 10µL final par séchage, et est mélangé à 10µL de matrice HCCA pour l'analyse en spectrométrie de masse MALDI-TOF (Figure 51).



**Figure 51 : Analyse par spectrométrie de masse MALDI-TOF des composés récoltés après séparation par HPLC provenant de l'extrait cellulaire réalisé sur un culot de *B. thuringiensis* HD-73 après 9H de culture en milieu LB à 30°C.**

Après analyse par spectrométrie de masse, la présence de pics de rapport  $m/z$  924,71, 946,71 et 962,70 peuvent être observés. Ces pics correspondent aux adduits  $H^+$ ,  $Na^+$  et  $K^+$  de la kurstakine C13 sous sa forme linéaire, indiquant la linéarisation partielle de la quantité de lipopeptide produit par *B. thuringiensis* HD-73 entre 8 et 9h de culture. Après 10h de culture, la kurstakine ne peut plus être détectée par HPLC.

L'étude de la production de kurstakine par *B. thuringiensis* HD-73 en fermenteur a permis de démontrer que la kurstakine était produite par *B. thuringiensis* peu après son entrée en phase stationnaire. De plus, la rapide diminution de la quantité de kurstakine mesurée par HPLC, ainsi que la modification du profil UV des molécules détectées d'une part, et des masses mesurées par spectrométrie de masse d'autre part, semble indiquer une rapide dégradation du lipopeptide peu de temps après sa synthèse. La linéarisation de la molécule pourrait ainsi représenter la première étape de la dégradation de la molécule.

Pour l'étude de la cinétique de production de kurstakine en milieu solide, la souche HD-73 a été cultivée en milieu AK agar et LB gélosé pendant 5 jours. L'extraction, la purification et

la quantification de kurstakine ont été réalisées selon le protocole décrit dans la partie matériel et méthodes (Figure 52).

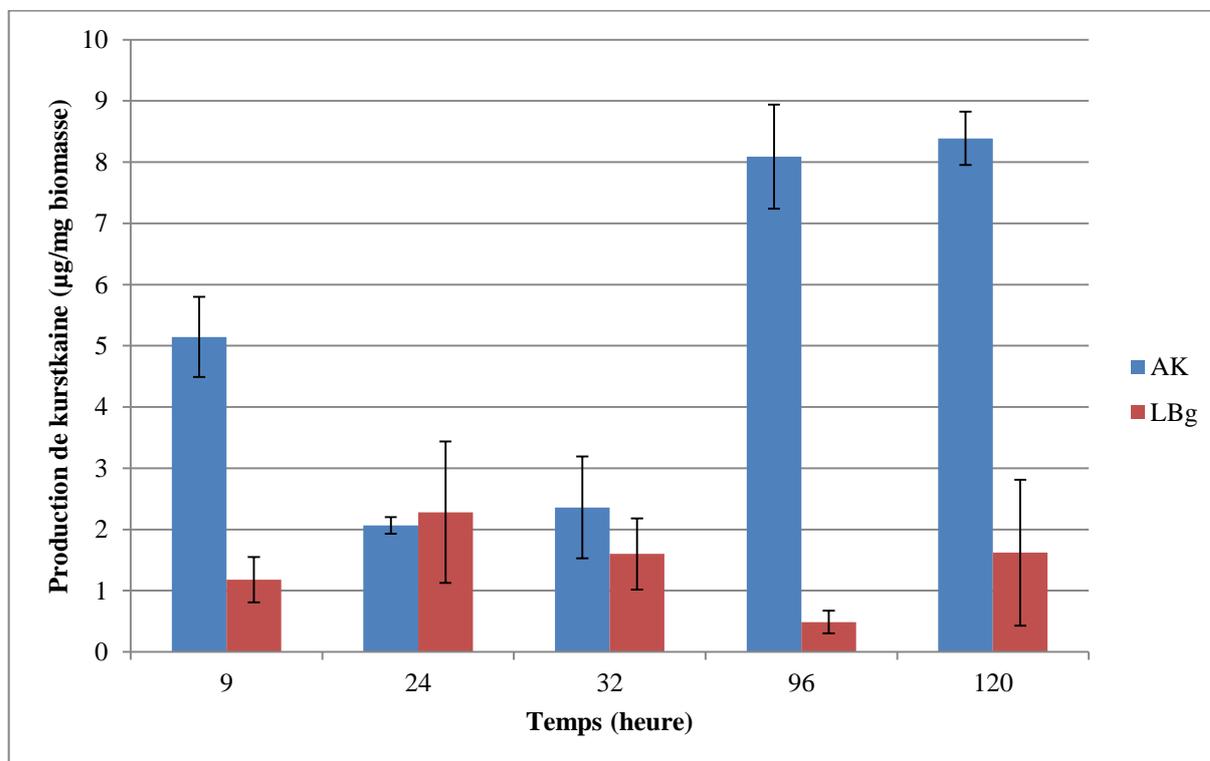


Figure 52 : Cinétique de production de kurstakine en milieu AK agar et LBG par la souche *Bacillus thuringiensis* HD-73.

En milieu de culture solide la production de kurstakine présente également de grande variation en fonction de l'âge de la culture. Ainsi, après 9h de culture, la production de kurstakine atteint 5 µg/mg de matière sèche en milieu AK agar. Après 24h et 32h de culture, le niveau de production mesuré est de 2 µg/mg de matière sèche environ. Enfin, le niveau de production à 96h et 120h est d'environ 8 µg/mg de matière sèche. Le maximum de production observé est donc atteint après 4 à 5 jours de cultures, soit le temps recommandé par le protocole décrit par Hathout *et al* (2000).

En milieu LBG, la production de kurstakine est d'environ 1 µg/mg de matière sèche après 9h de culture, et atteint un maximum après 24h de culture, avec 2 µg/mg de matière sèche. La quantité de kurstakine diminue ensuite jusqu'à 0,5 µg/mg de matière sèche après 96h, puis s'établit à 2 µg/mg de matière sèche après 120h. Néanmoins, les imprécisions de mesures liées aux faibles quantités de molécule détectée ne permettent pas d'établir de différence significative des mesures entre les différents temps de culture pour ce milieu.

## Discussion

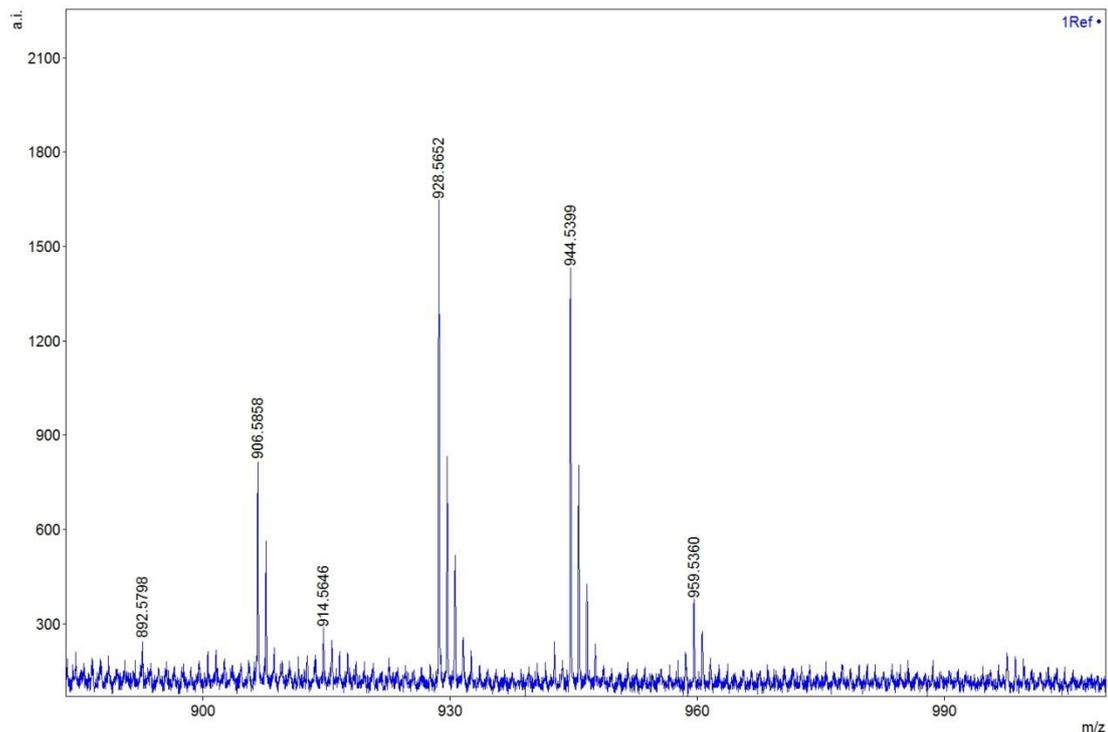
La cinétique de production de kurstakine en milieu liquide correspond à la production d'un métabolite secondaire, la production étant impossible à évaluer avant la phase stationnaire. Le niveau de production maximum mesuré avec les deux milieux testés se situe après 24h de culture, diminuant ensuite à 72h de culture. Cette diminution de la quantité de lipopeptide pourrait alors correspondre à un recyclage des acides aminés composant la kurstakine par l'action d'une enzyme permettant cette dégradation. La production est plus importante en milieu AK liquide qu'en milieu LB, avec un taux de production 5 fois supérieur avec le premier milieu après 24h de culture, indiquant que ce milieu favorise la production de la molécule.

De plus, l'analyse des échantillons issus de la production du lipopeptide en fermenteur permet d'identifier la présence de formes linéaires du lipopeptide une heure après leur production. La linéarisation de la molécule pourrait représenter la première étape de la dégradation observée précédemment.

En milieu solide, la cinétique de production est plus complexe. En milieu LB gélosé, le niveau de production varie peu au cours des 5 jours de culture. Le taux de production connaît un maximum d'environ 2 µg/mg de matière sèche après 24 h et 120 h. En milieu AK agar, la quantité de kurstakine produite est plus importante, et atteint son maximum après 96h de culture, avec environ 8µg/mg de matière sèche. Ces résultats suggèrent une production tardive du lipopeptide en milieu solide AK agar.

### *1.2.5 Etude de la relation entre la sporulation de Bt et la production de kurstakine*

La production de kurstakine a été, depuis la découverte de cette molécule, associée au phénomène de sporulation de *B. thuringiensis*. Cela est notamment dû à l'état physiologique des cellules utilisées lors de la découverte du lipopeptide. Afin d'en savoir plus sur la relation entre la production de kurstakine et la sporulation des cellules, la capacité de production du lipopeptide par la souche Bt 407Δspo0A a été étudiée. Cette souche dont le gène spo0A, codant pour un activateur des gènes liés à la sporulation, a été interrompu est donc incapable de sporuler. Pour cela, une analyse par spectrométrie de masse MALDI-TOF a été réalisée sur une colonie cultivée sur milieu LB gélosé pendant 24 heures à 37°C (Figure 53).



**Figure 53 : Analyse par spectrométrie de masse d'une colonie de la souche Bt407 $\Delta$ spo0A cultivée sur milieu LB gélosé pendant 24 heures à 37°C.**

L'analyse par spectrométrie de masse MALDI-TOF permet d'identifier les adduits  $H^+$ ,  $Na^+$  et  $K^+$  de la kurstakine C12 et C13. Les pics correspondants aux formes C11 et C14 ne sont pas détectés. Il est néanmoins difficile de déterminer s'il s'agit d'une réelle absence de production de ces deux formes, ou de l'effet de la limite de détection par l'appareil.

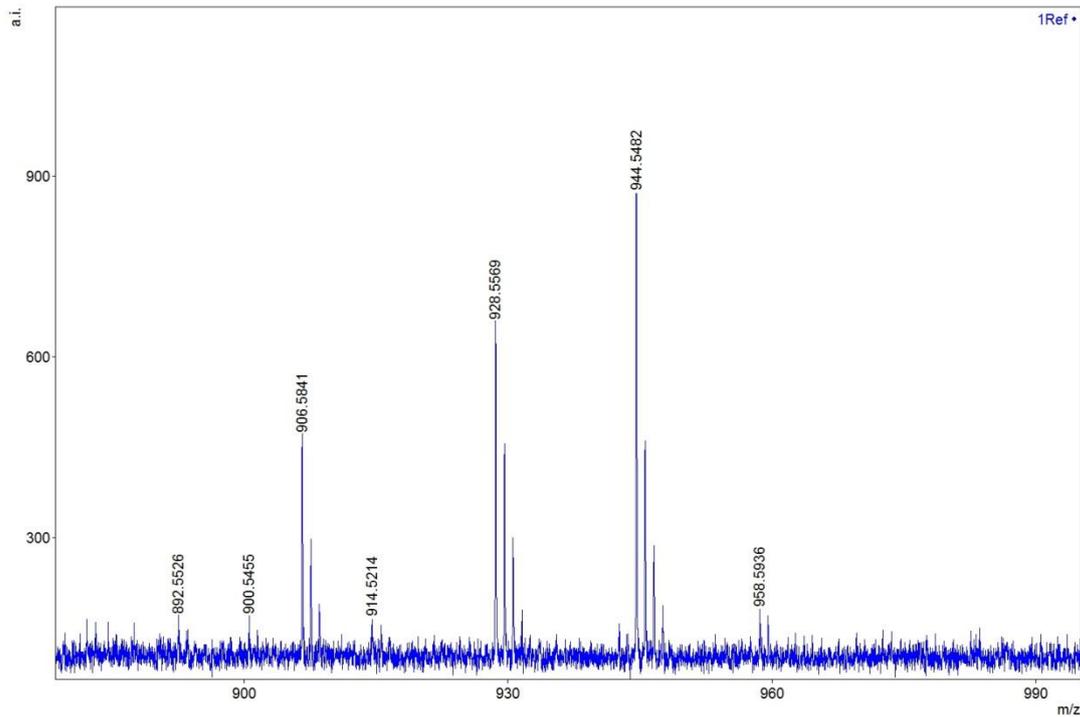
Les résultats démontrent que la souche Bt407 $\Delta$ spo0A produit de la kurstakine malgré la délétion de l'activateur transcriptionnel principal de la sporulation chez *B. thuringiensis*. Cela indique que la sporulation des cellules de Bt n'est pas nécessaire à la production de kurstakine.

### *1.2.6 Effet de l'inactivation de la production de toxines de B. thuringiensis sur la production de kurstakine*

La production de toxines protéiques sous forme de cristaux est l'une des particularités métaboliques les plus importantes de *B. thuringiensis*. La production de ces cristaux pouvant présenter pour la souche une importante utilisation de nutriments ainsi qu'une utilisation importante d'énergie, la suppression de cette activité de synthèse pourrait donc avoir un effet bénéfique sur la production de kurstakine.

L'effet du curage du plasmide portant les gènes permettant la production de cristaux protéiques toxiques par les souches de Bt a donc été étudié au moyen de deux souches modifiées : Bt 407 cry- et Bt HD-73 cry-.

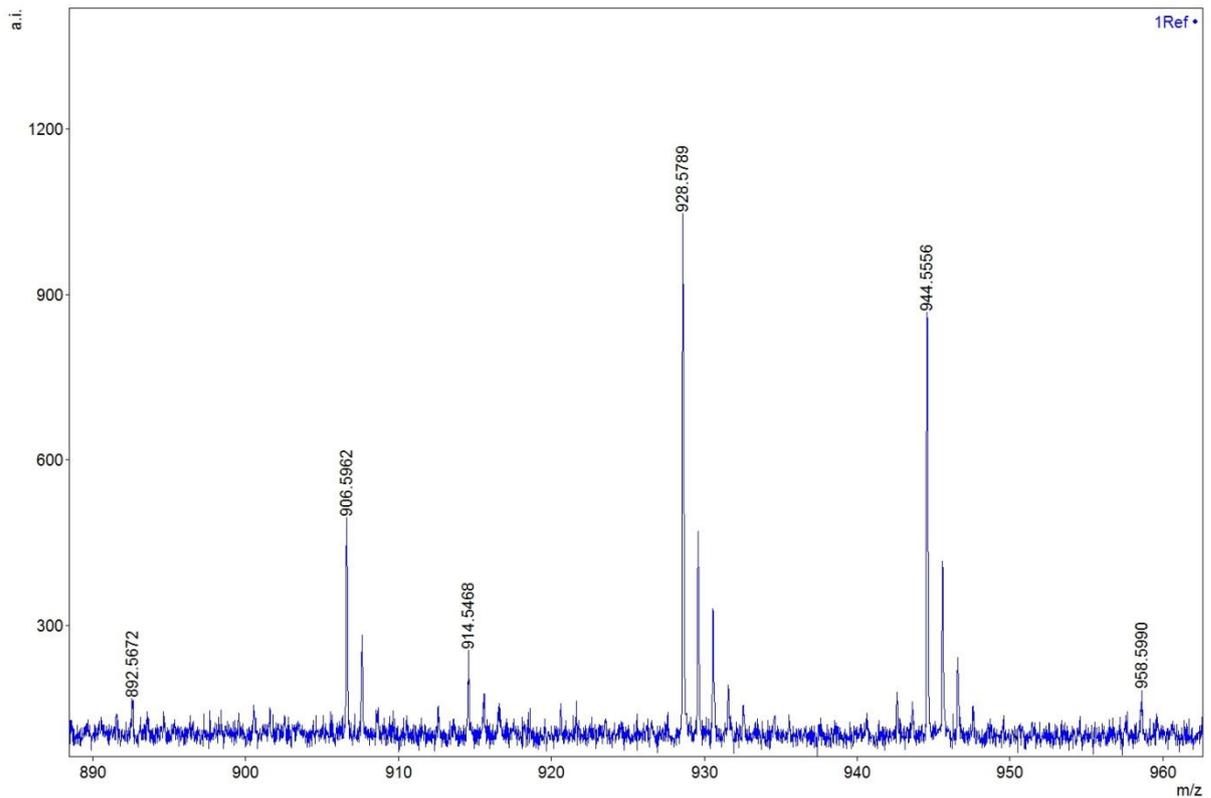
Dans un premier temps, l'analyse des produits des différentes souches a été réalisée par spectrométrie de masse MALDI-TOF à partir de colonies cultivées sur milieu LB gélosé pendant 24 heures à 30°C (Figure 54).



**Figure 54 : Spectre de masse MALDI-TOF réalisé sur une colonie de la souche Bt407Cry+ cultivée sur milieu LB gélosé pendant 24 heures à 37°C.**

La souche Bt 407 sauvage présente les pics des adduits des kurstakines en C11, C12, C13 et C14.

L'analyse a été réalisée dans les mêmes conditions pour la souche Bt407 cry- chez laquelle le plasmide portant les gènes cry a été supprimé (Figure 55).

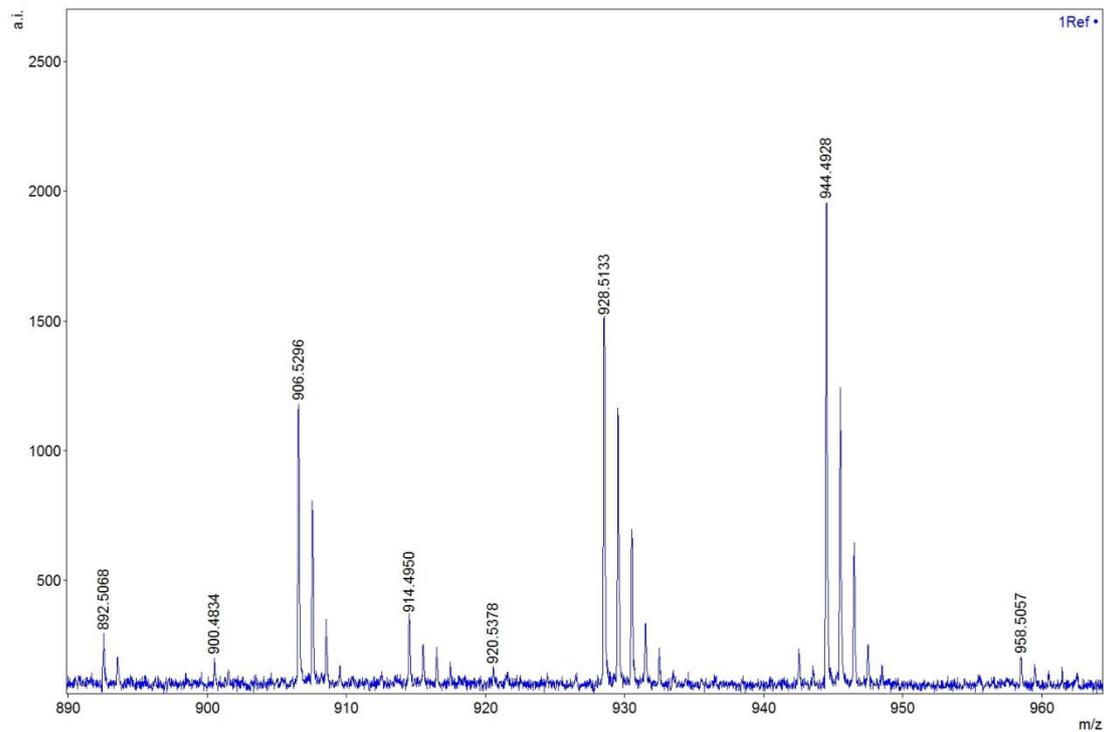


**Figure 55 : Spectre de masse MALDI-TOF réalisé sur une colonie de la souche Bt407Cry- cultivée sur milieu LB gélosé pendant 24 heures à 37°C.**

Les kurstakines de forme C11 à C14 sont retrouvées sur les cellules de la souche Bt407 cry-.

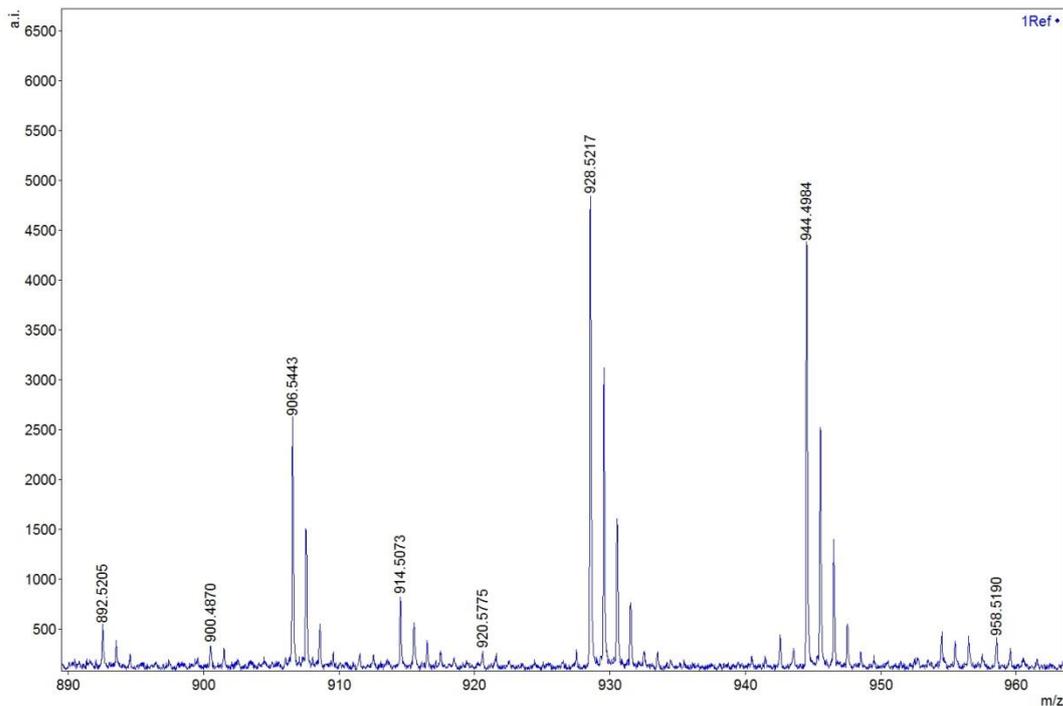
Les souches Bt407 sauvage et Bt 407 cry- produisent donc les mêmes formes de kurstakine, indiquant que la suppression du plasmide portant les gènes responsables de la production de toxines n'entraîne pas de modification qualitative des formes de kursakine produites.

Afin de confirmer ce résultat, les produits de la souche sauvage *B. thuringiensis* HD-73 sauvage et de la souche modifiée *B. thuringiensis* HD-73 cry- chez laquelle le plasmide portant les gènes de production de toxine a été supprimé ont été analysés dans les mêmes conditions.



**Figure 56 : Spectre de masse MALDI-TOF réalisé sur une colonie de la souche *B. thuringiensis* HD-73 cultivée sur milieu LB gélosé pendant 24 heures à 37°C.**

L'analyse par spectrométrie de masse MALDI-TOF de la souche *B. thuringiensis* HD-73 permet de détecter les formes C11 à C14 de la kurstakine. L'analyse similaire de la souche *B. thuringiensis* HD-73 chez laquelle le plasmide portant les gènes *cry* a été délété est présentée dans la figure suivante.

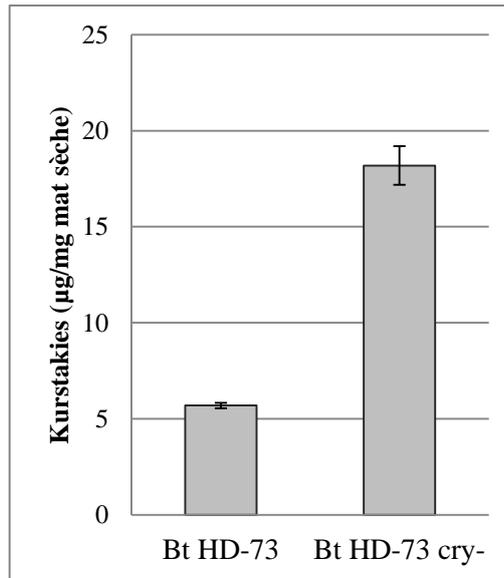


**Figure 57 :** Spectre de masse MALDI-TOF réalisé sur une colonie de la souche *B. thuringiensis* HD-73 cry- cultivée sur milieu LB gélosé pendant 24 heures à 37°C.

L'analyse réalisée sur la souche *B. thuringiensis* HD-73 cry- permet d'identifier les mêmes formes du lipopeptide que la souche non modifiée.

Le curage du plasmide portant les gènes permettant la production de cristaux protéiques toxiques n'a pas d'effet majeur sur la variété de kurstakine produite par les souches de Bt, les mêmes formes de la molécule étant retrouvées dans la souche sauvage et la souche curée du plasmide.

L'effet du curage du plasmide portant les gènes de production de toxine sur la production de kurstakine par *B. thuringiensis* a également été analysé quantitativement par un dosage de la production du lipopeptide par HPLC. Ce dosage a été réalisé à partir d'extrait cellulaires réalisés selon le protocole décrit dans la partie matériel et méthodes après culture des souches sur milieu AK agar pendant 1 nuit à 37°C puis 3 jours à température ambiante (Figure 58).



**Figure 58 : Effet du curage du plasmide portant les gènes cry chez la souche Bt HD-73 sur la production de kurstakine. Le dosage HPLC a été réalisé à partir d'extraits cellulaires de cellules cultivées sur milieu AK agar pendant 1 nuit à 37°C et 3 jours à température ambiante et purifiée sur SPE C8.**

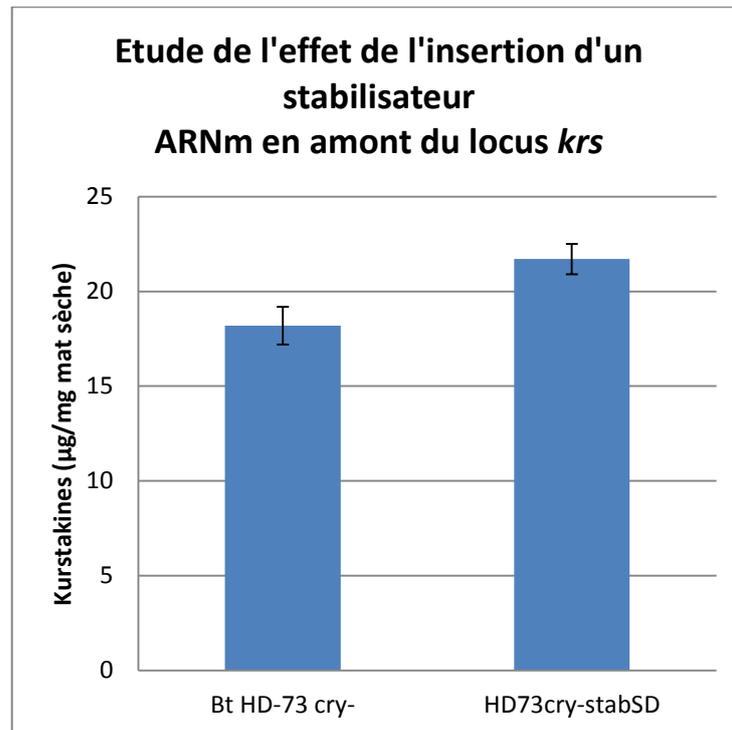
La souche sauvage Bt HD-73 présente un niveau de production de kurstakine d'environ 6 µg/mg de matière sèche. Après curage du plasmide, la souche produit environ 18 µg/mg de matière sèche. Pour la souche Bt 407, le curage du plasmide permet l'augmentation de la production de 1 à 5 µg/mg de matière sèche.

Le curage du plasmide a ainsi un effet positif sur la production de kurstakine d'un point de vue quantitatif, avec une augmentation du taux de production de lipopeptide ramené à la matière sèche d'un facteur 3 à 5 en fonction de la souche. Cette augmentation de la production pourrait être expliquée par l'augmentation de la disponibilité des ressources de la souche d'un point de vue énergétique et du point de vue des précurseurs non mobilisés pour la production de cristaux..

### *1.2.7 Effet de l'augmentation de la stabilité de l'ARNm de la kurstakine synthétase sur la production de kurstakine*

La stabilité des ARN messagers issus des gènes de production est essentielle à la production de la synthétase responsable de la production de la kurstakine. Un problème de stabilité des ARNm entrainerait une synthèse très faible du complexe enzymatique produisant le lipopeptide. L'importance de la stabilité de l'ARNm des gènes *krsEABC* a donc été évaluée ici par le dosage de la kurstakine par les souches Bt HD-73 cry- STAB-SD et Bt HD-73 cry-spo0A STAB-SD... chez lesquelles la séquence STAB-SD a été intégrée en amont du gène *krsE*. Cette séquence de Shine-Dalgarno a été identifiée chez *B. thuringiensis* comme agissant

en tant que stabilisateur d'ARNm du gène de toxine cryIII<sub>A</sub>. L'insertion de cette séquence en amont de *krsE* permet donc d'augmenter la stabilité de l'ARNm de la kurstakine synthétase. Les différentes souches modifiées ont été cultivées sur milieu AK agar et utilisées pour la préparation d'extraits cellulaires selon la méthode décrite dans la partie matériels et méthodes. Les résultats de production de ces souches et des souches mère correspondantes sont présentés dans la Figure 59.



**Figure 59 : Taux de production de kurstakine sur milieu AK agar après 1 nuit à 37°C et 3 jours à température ambiante chez les souches Bt HD-73 cry- et Bt HD-73 cry-STAB-SD.**

La souche Bt HD-73 cry- produit 18,2 µg de kurstakine par mg de matière sèche. L'introduction du stabilisateur STAB-SD chez la souche entraîne une augmentation de la production à 21,7 µg/mg, soit une augmentation d'environ 20%.

L'effet de l'introduction du stabilisateur d'ARNm en amont du locus *krs* chez *B. thuringiensis* entraîne donc une légère augmentation de la production chez la souche HD73cry-. En revanche, cet effet semble proportionnellement plus important chez la souche HD73cry- spo0A.

### 1.2.8 Rôle du gène *krsE* dans le transport de la kurstakine

Le rôle du produit du gène *krsE*, dont l'analyse bioinformatique démontre une proximité avec la famille des protéines de transport de type Major Facilitator Superfamily (MFS-1), est à ce jour inconnu. Néanmoins, la prédiction de son activité semble indiquer l'implication de la protéine KrsE dans le transport de la kurstakine. Afin de déterminer le rôle joué par cette protéine dans la synthèse de la kurstakine, l'effet de la délétion et de la surexpression du gène *krsE* sur la production de kurstakine a été étudié lors d'une culture en milieu liquide. Une souche Bt HD73 $\Delta$ *krsE* *krs* :: STAB-SD a été utilisée pour analyser l'effet de la délétion du gène de la protéine KrsE sur la production et le transport de la kurstakine. Une souche HD73 *krs* :: STAB-SD pHT1618T-pxyl $\Omega$ *krsE* a été utilisée pour l'évaluation de l'effet de la surexpression du gène *krsE* sur la production et le transport de kurstakine. Cette souche, chez laquelle le promoteur de la kurstakine synthétase est inductible par l'addition de xylose, a été cultivée en milieu LB liquide, en absence ou en présence de xylose à 20 mM (LBxyl), permettant l'induction de la transcription du gène *krsE*.

Les cultures liquides ont été réalisées dans 100 mL de milieu LB ou LBxyl en erlenmeyer de 500 mL, à 30°C pendant 24 heures. Un prélèvement de 5 mL a été effectué toutes les heures pendant les 8 premières heures de culture. Un prélèvement de 5 mL a également été effectué après 24, 28 et 32 heures de culture. Pour chaque prélèvement, le culot et le surnageant ont été séparés par centrifugation à 5000g pendant 5 minutes, et l'extraction, la purification et l'analyse de la production de kurstakine ont été réalisés comme décrit dans la partie matériel et méthodes. Les courbes de croissance de la souche souche HD73 *krs* :: STAB-SD pHT1618T-pxyl $\Omega$ *krsE* cultivée en milieu LB en absence ou en présence de xylose à 20mM sont présentés dans la Figure 60.

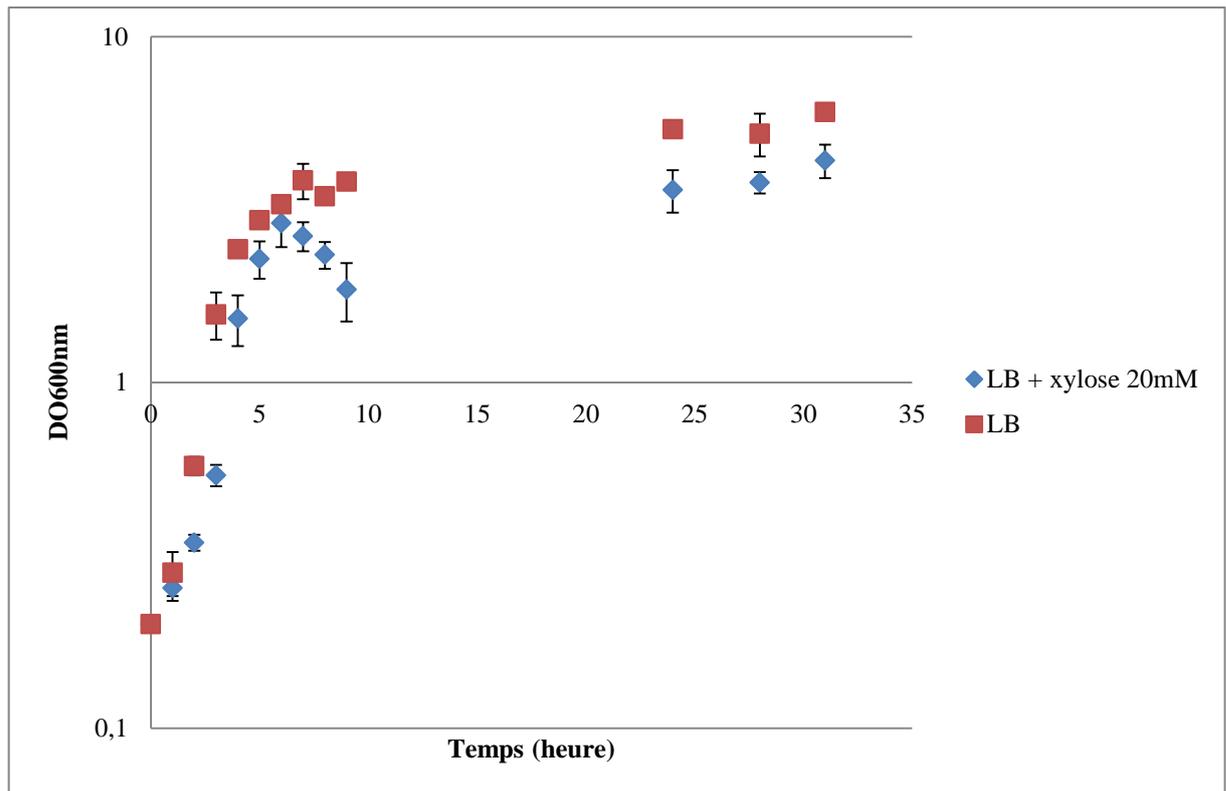
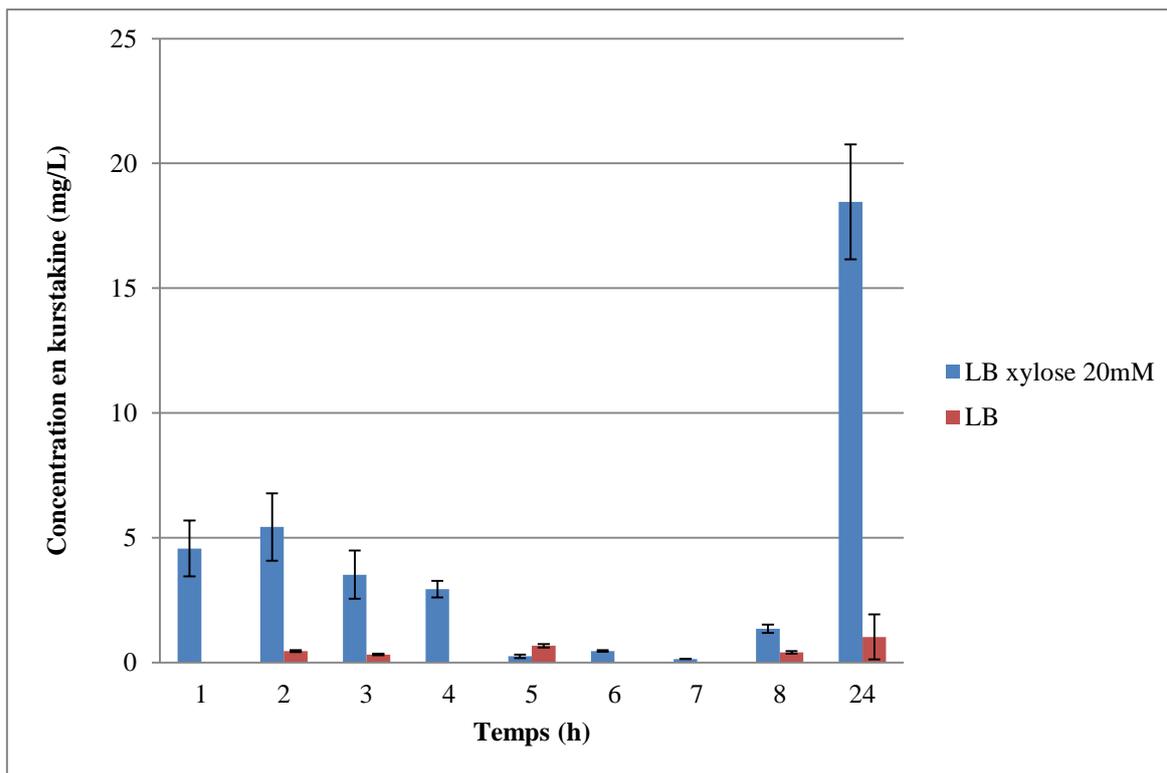


Figure 60 : Courbe de croissance de la souche *B. thuringiensis* HD73 krs :: STAB-SD pHT1618T-pxylQkrsE en milieu LB liquide avec ou sans addition de xylose à 20 mM.

En milieu LB, la souche Bt HD73 krs :: STAB-SD pHT1618T-pxylQkrsE présente une phase exponentielle de croissance d'environ 3 heures. La souche entre ensuite en phase de ralentissement, jusqu'au début de la phase stationnaire, à environ 10 heures de culture, atteignant une DO<sub>600nm</sub> maximale de 6,0 après 32 heures de culture.

En milieu LBxyl, la souche présente le même profil de croissance au niveau de la phase exponentielle. Néanmoins, là où la souche présente une simple phase de ralentissement en milieu LB, la croissance en présence de xylose voit la DO<sub>600nm</sub> de la culture chuter à partir de 6 heures de cultures, et ce jusqu'à 10 heures, passant de 2,9 à 1,8. A 24 heures de culture, la souche cultivée en présence de xylose présente une DO<sub>600nm</sub> plus faible qu'en milieu LB.

La suivi de la concentration de kurstakine présente dans le surnageant de culture de la souche dans les différentes conditions est présentée dans la Figure 61.



**Figure 61 : Dosage de kurstakine dans les surnageants de culture de la souche *B. thuringiensis* HD73 krs :: STAB-SD pHT1618T-pxylQkrsE cultivée en milieu LB et LBxyl. La concentration du lipopeptide, évaluée par dosage HPLC, est exprimée en mg/L.**

En milieu LB, la concentration en kurstakine dans le surnageant de culture est relativement faible, voire impossible à évaluer par HPLC, étant inférieure au seuil de détection de l'appareil malgré l'étape de concentration réalisée sur colonne SPE. La concentration maximale mesurée est obtenue après 24 heures de cultures, avec 1,02 mg/L de kurstakine dans le surnageant de culture.

En milieu LBxyl, les concentrations mesurées dès les premières heures de cultures sont de l'ordre de 4,6 et 5,4 mg/L pour la première et la deuxième heure de culture respectivement. La concentration diminue ensuite progressivement jusqu'à la 7<sup>ème</sup> heure de culture, avec une concentration de 0,1 mg/L de lipopeptide. A 8 heures de culture, la concentration mesurée augmente à 1,3 mg/L. Après 24 heures de culture, la concentration mesurée dans le surnageant de culture de la souche cultivée en LBxyl est de 18,46 mg/L.

## Discussion

Jusqu'à présent, le rôle joué par le gène *krsE* dans le mécanisme de production de kurstakine n'avait pas été élucidé. Si l'analyse de la séquence du gène permettait d'associer la protéine produite à la famille de transporteurs MSF-1, la démonstration expérimentale du rôle joué par celle-ci n'avait pas été réalisée. La découverte de la kurstakine ayant été réalisée en utilisant directement des spores de Bt, et l'absence de dosage du lipopeptide dans le milieu de culture a longtemps amené à penser que celui-ci, à l'inverse des autres lipopeptides de *Bacillus* étudiés, tels la surfactine, la fengycine ou la mycosubtilline, n'était pas excrétée dans le milieu de culture liquide par les bactéries.

La possibilité de doser les kurstakines dans le milieu de culture de la souche HD73 *krs* :: STAB-SD pHT1618T-pxyl $\Omega$ krsE cultivée dans un milieu permettant l'induction de la production de KrsE permet de confirmer le rôle joué par cette protéine dans le transport du lipopeptide dans le milieu de culture. De plus, la faible quantité du lipopeptide dosé dans le cas où la souche est cultivée sans inducteur peut n'être uniquement due qu'à la fuite transcriptionnelle du gène *krsE*. Ainsi, il paraît probable que la kurstakine n'est excrétée dans le milieu de culture qu'à condition d'une transcription du gène *krsE*.

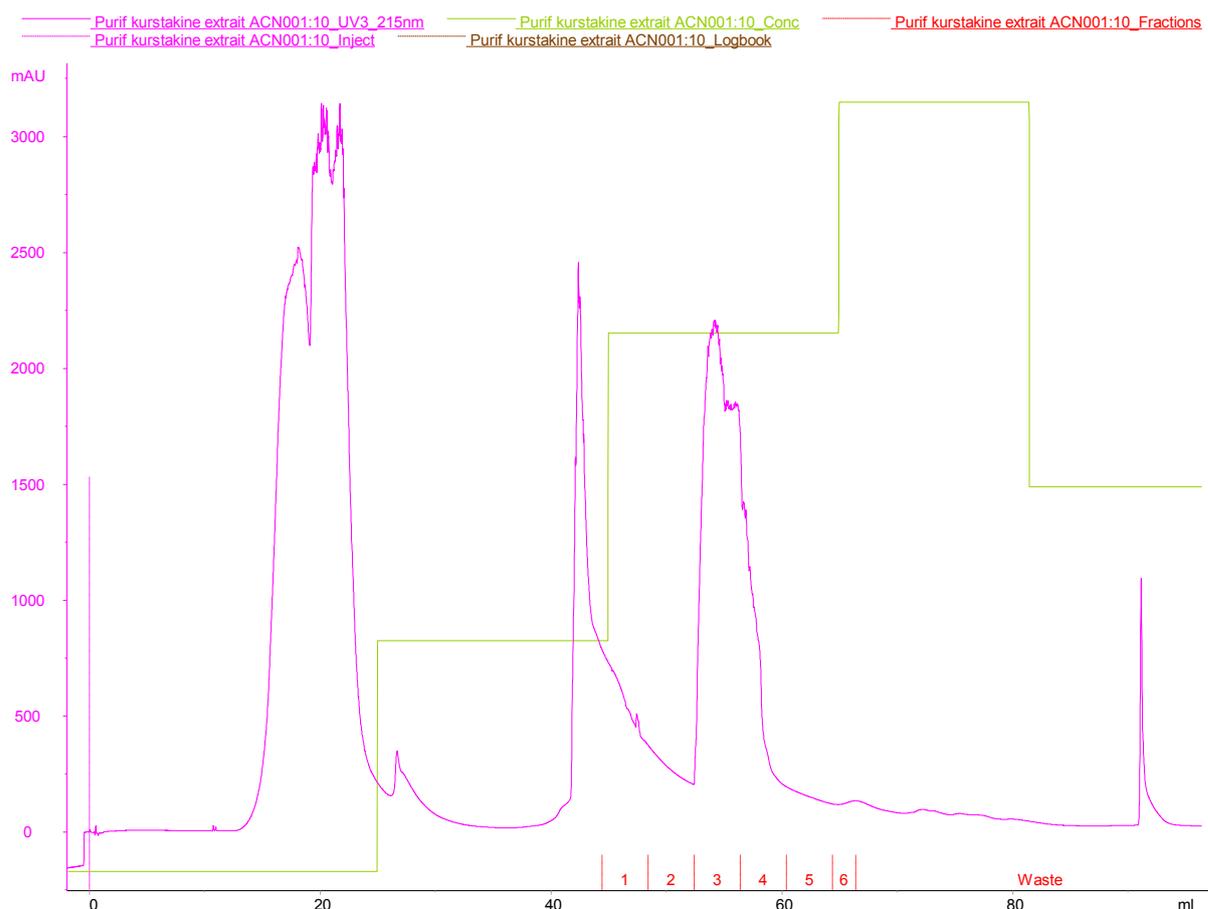
Il est surprenant de trouver des concentrations de 4 à 5 mg/L de kurstakine dans les surnageants de culture de la souche *B. thuringiensis* HD73 *krs* :: STAB-SD pHT1618T-pxyl $\Omega$ krsE dès les premières heures de culture lorsque la souche est cultivée avec l'inducteur. Cette concentration observée, qui diminue après 2 heures de culture, peut néanmoins être due au passage des cellules de la préculture, réalisée en milieu LB, à un milieu de culture LBxyl, entraînant alors l'excrétion de la kurstakine présente dans les cellules de la préculture. En revanche, la raison de la diminution de la quantité de lipopeptide pendant la suite de la phase exponentielle (de 2 à 3 heures) et pendant la phase de ralentissement (de 3 à 7 heures) n'a pas été élucidée. Il est possible que les lipopeptides aient été dégradés par la souche.

La multiplication d'un facteur 20 environ de la quantité de kurstakine dosée dans le surnageant de culture par l'induction de la transcription du gène *krsE* permet d'envisager le potentiel d'une telle modification. En effet, la production et la purification de lipopeptides par des souches bactériennes est bien plus facile à réaliser au moyen de cultures liquides.

## **Purification de kurstakine par FPLC**

A ce jour, aucun procédé de purification de la kurstakine à l'échelle semi-préparative n'est en place pour la purification de kurstakine au sein du laboratoire. De plus, les faibles concentrations obtenues pour la production de ce lipopeptide ne permettent pas l'utilisation des techniques actuellement mises en oeuvre au sein de la structure pour la purification d'autres lipopeptides de *Bacillus*. En effet, les lipopeptides comme la surfactine ou la mycosubtiline sont actuellement produits au laboratoire au moyen de culture en fermenteur, et les volumes de surnageants utilisés pour la purification sont traités par microfiltration et ultrafiltration. Cette dernière technique est basée sur la formation de micelles par les lipopeptides pour des concentrations supérieures à leur concentration micellaire critique. La concentration micellaire critique (CMC) de la kurstakine n'ayant pas été mesurée, notamment en raison de la difficulté à le produire et le purifier en quantité suffisante, cette technique n'est pas aujourd'hui transférable à la purification de la kurstakine.

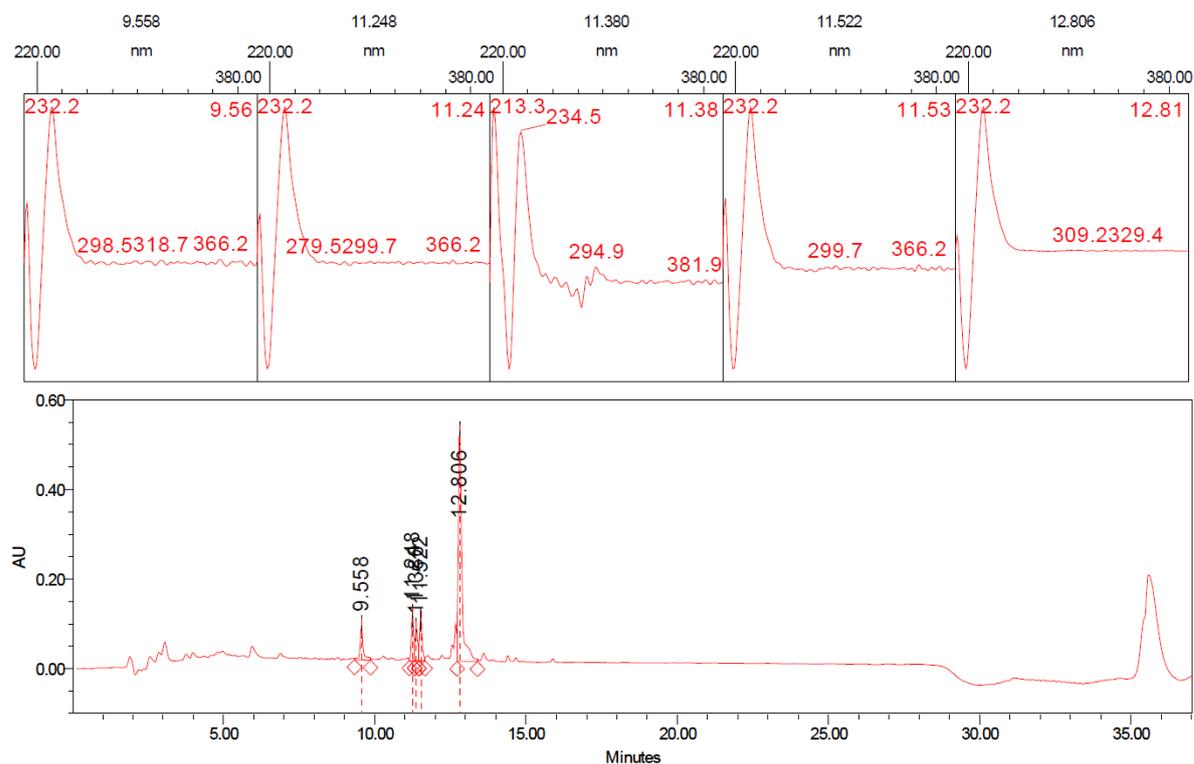
Une purification de la kurstakine par FPLC a donc été réalisée sur 20 mL d'extrait purifié en kurstakine préparé à partir de 50 boîtes de *B. thuringiensis* HD-1. Après l'injection de l'extrait sur la colonne FPLC chargée en phase greffée en C8, le suivi de l'élution du lipopeptide est réalisé par la mesure de l'absorbance à 215nm (Figure 62).



**Figure 62 : Chromatogramme d'extrait de kurstakine analysé par FPLC. Le suivi de l'absorbance à 215nm est représenté par le graphique rose. Le gradient de concentration en acétonitrile est représenté en vert. Les fractions prélevées en sortie de colonne sont représentées en rouge.**

Le chromatogramme d'absorption à 215 nm présente plusieurs pics ou ensembles de pics. Un pic large est observé entre 10 et 30 minutes d'élution. Il correspond aux composants de l'extrait pouvant être élués par le solvant H<sub>2</sub>O/TFA 0.1%. Un deuxième pic est observé entre 40 et 50 minutes d'élution, correspondant aux composants de l'extrait nettoyés de la colonne par le passage de 30% d'acétonitrile. Un troisième pic est observé entre 50 et 60 minutes d'élution, correspondant à l'élution des composés par le passage du solvant composé à 70% d'acétonitrile. De 45 à 65 minutes, des fractions sont collectées à la sortie du circuit afin d'identifier la nature des molécules présentes dans le pic compris entre 50 et 60 minutes.

Une fois concentrées sous vide, les fractions sont reprises dans un volume plus faible d'ACN/H<sub>2</sub>O TFA, et analysées en HPLC pour la détection de kurstakine. L'analyse de la fraction 3 contenant potentiellement le plus de lipopeptide est présenté ci-dessous (Figure 63).



**Figure 63 : Chromatogramme de la fraction 3 récoltée lors de la purification de kurstakine par FPLC.**

Le profil HPLC de l'échantillon collecté à la suite de la purification par FPLC présente les 4 pics caractéristiques de la kurstakine aux temps de rétention compris entre 9 et 13 minutes. Le chromatogramme permet également d'évaluer la qualité de la purification. En effet, les pics de kurstakines représentent la majorité des pics présents sur le chromatogramme. Néanmoins, cette technique de purification n'a pas permis de récolter une quantité suffisante en lipopeptide permettant d'effectuer la pesée du produit sec, nécessaire au calcul précis du degré de pureté du produit récupéré.

## Discussion

La purification de la kurstakine par FPLC a pu être réalisée au moyen d'une colonne FPLC remplie de silice greffée en phase C8, en adaptant le programme d'élution utilisé en HPLC pour séparer la kurstakine des autres composés de l'extrait cellulaire. La comparaison du chromatogramme HPLC réalisé à l'issue de la purification par FPLC avec les chromatogrammes obtenus après purification par colonne SPE permet d'obtenir une meilleure qualité de purification à partir d'extraits cellulaires de Bt. La mise au point de la purification de kurstakine par FPLC pourra permettre à terme la purification directe de surnageants de culture de Bt pour la purification de kurstakine, en ajustant la capacité de la colonne aux besoins.

### 1.3 Tests d'activités biologiques de la kurstakine

Comme décrit dans la partie synthèse bibliographique, les lipopeptides de *Bacillus* possèdent une variété importante d'activités industriellement valorisables (antibactérienne, antifongique, surfactant...). A ce jour, une seule activité antifongique a été démontrée chez la kurstakine, contre la seule souche *Stachybrotris chartarum*. Le but de cette partie est d'évaluer les activités biologiques potentiellement valorisables de la kurstakine, notamment en tant qu'antifongique et antibactérien.

#### 1.3.1 Test d'activité antifongique

Afin de déterminer le potentiel d'utilisation de la kurstakine en phytoprotection, des tests d'activité antifongique sont réalisés à partir d'extraits purifiés en kurstakine. Chaque souche de moisissure sur laquelle l'effet de la kurstakine est testé est cultivée en milieu PDA à partir du centre de la boîte de Pétri, et l'inoculum est entouré de disques contenant respectivement 25, 50 et 100µg de kurstakine. L'activité antifongique est évaluée par l'inhibition de la croissance des moisissures testées autour des disques contenant le lipopeptide.

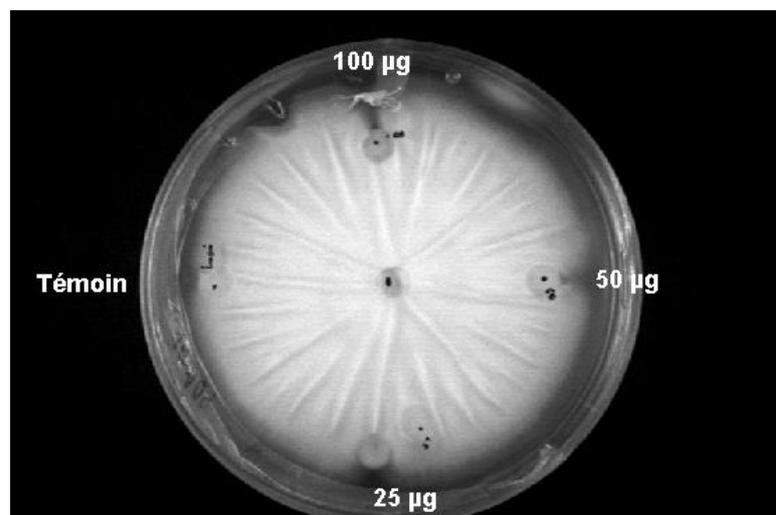
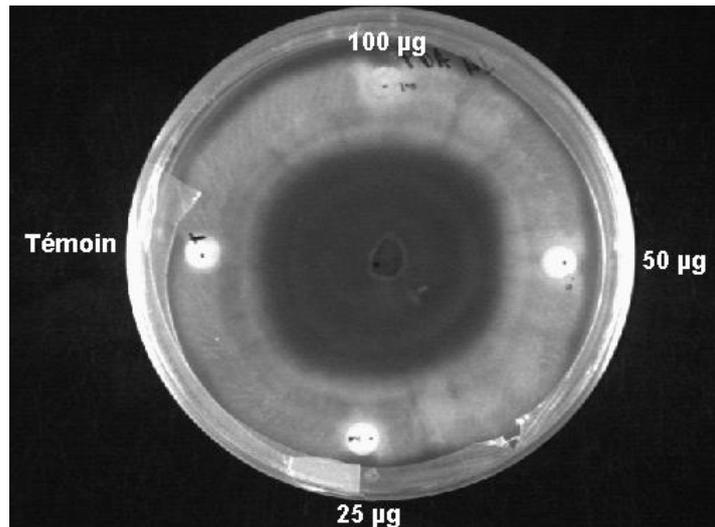


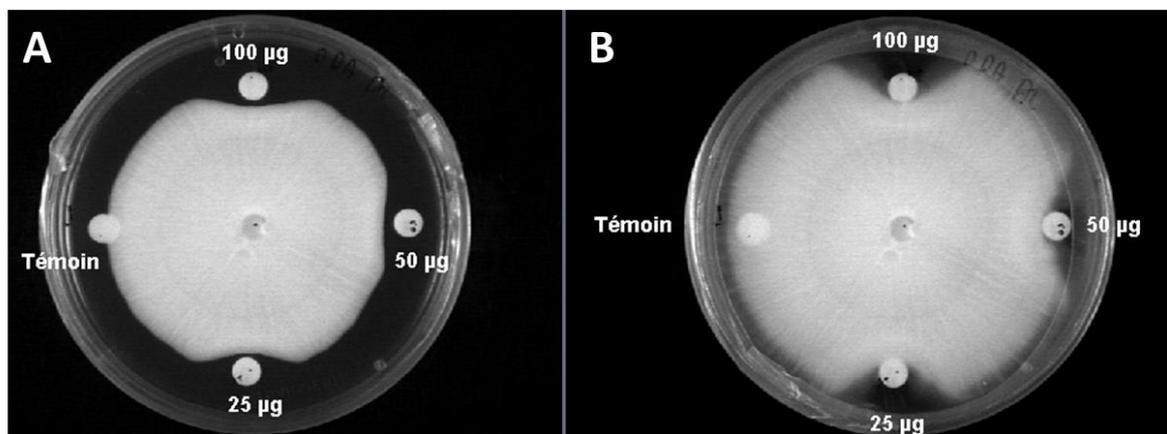
Figure 64 : Test d'activité antifongique d'extraits purifiés de kurstakine contre la souche *Rhizoctonia solani*. Les quantités indiquées correspondent à la quantité de lipopeptide présente sur le disque. Le témoin négatif est composé d'acétonitrile à 30%.

La souche *Rhizoctonia solanii* a pu se développer sur toute la surface de milieu gélosé PDA, en recouvrant également les disques imbibés d'extrait de lipopeptide (Figure 64). La kurstakine n'a donc aucun effet antagoniste sur la croissance de la souche.



**Figure 65 :** Test d'activité antifongique d'extraits purifiés de kurstakine contre la souche *Fusarium oxysporum*. Les quantités indiquées correspondent à la quantité de lipopeptide présente sur le disque. Le témoin négatif est composé d'acétonitrile à 30%.

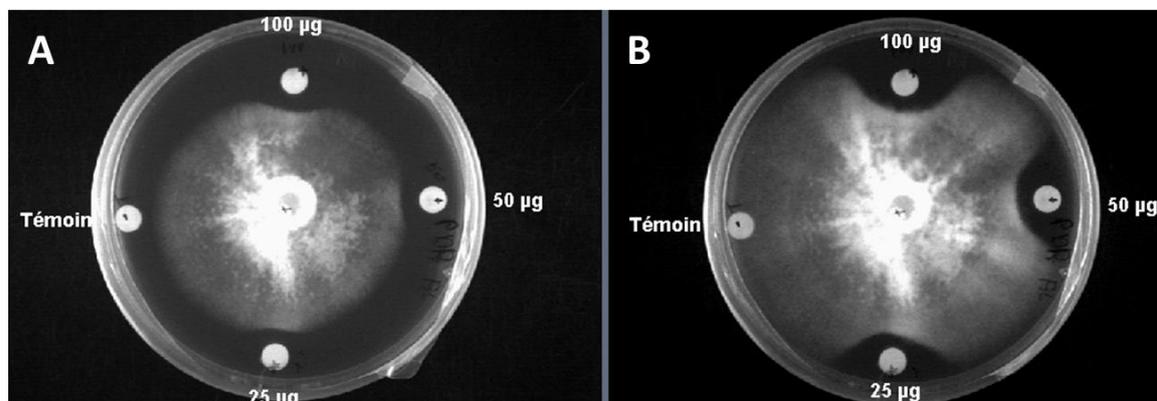
Comme pour *Rhizoctonia solanii*, la souche *Fusarium oxysporum* a pu croître sur toute la surface de la boîte de Pétri et a recouvert les disques d'extrait de kurstakine (Figure 65). Celle-ci n'a donc pas d'effet antagoniste contre *F. oxysporum* aux concentrations testées.



**Figure 66 :** Test d'activité antifongique d'extraits purifiés de kurstakine contre la souche *Galactomyces geotrichum* après 3 jours (A) et 5 jours d'incubation (B). Les quantités indiquées correspondent à la quantité de lipopeptide présente sur le disque. Le témoin négatif est composé d'acétonitrile à 30%.

La souche *Galactomyces geotrichum* n'a pas colonisé toute la surface de la boîte de Pétri. Trois jours après le dépôt de la moisissure, sa croissance a atteint le disque témoin et a commencé à le recouvrir. En revanche, un halo d'inhibition de croissance est visible autour de tous les disques contenant de la kurstakine. Après 5 jours de culture, la souche a totalement recouvert le disque témoin, mais n'a pas recouvert les disques contenant l'extrait de lipopeptide. De plus, une inhibition de croissance est visible au niveau des disques, vers

l'extérieur de la boîte de Pétri (Figure 66). La kurstakine possède donc un effet antifongique sur la souche *Galactomyces geotrichum*.



**Figure 67 :** Test d'activité antifongique d'extraits purifiés de kurstakine contre la souche *Botrytis cinerea* après 5 jours (A) et 7 jours d'incubation (B). Les quantités indiquées correspondent à la quantité de lipopeptide présente sur le disque. Le témoin négatif est composé d'acétonitrile à 30%.

La kurstakine est aussi active contre la souche *Botrytis cinerea*. En effet, après 5 jours de culture, un halo d'inhibition est visible autour des disques contenant de la kurstakine. Après 7 jours de culture, la souche a totalement recouvert le disque témoin, tandis qu'un halo d'inhibition de croissance d'un diamètre d'environ 1,5cm est visible autour des disques contenant du lipopeptide, et ce à toutes les concentrations testées (Figure 67).

## Discussion

Sur les 4 souches testées, la kurstakine présente donc une activité antifongique contre 2 d'entre elles, à savoir *Galactomyces geotrichum*, un agent de la pourriture aigre touchant notamment la tomate et contre *Botrytis cinerea*, un agent de la pourriture grise touchant plusieurs centaines d'espèces végétales. Jusqu'à présent, une activité antifongique de la kurstakine n'avait été identifiée que contre la souche *Stachybotrys chartarum*. La découverte d'activité antifongique du lipopeptide contre des souches phytopathogènes permet d'imaginer la possibilité d'utiliser cette molécule dans le domaine de la phytoprotection contre ces champignons.

### 1.3.2 Tests d'activité antibactérienne et antifongique en plaque (CMI)

Afin d'étudier le potentiel antibactérien de la kurstakine, des tests d'activités antibactériennes ont été réalisées contre les souches *Escherichia coli* ATCC25922 et *Candida albicans* ATCC10231 selon le protocole décrit dans la partie matériel et méthode.

Les souches *Escherichia coli* ATCC25922 et *Candida albicans* ATCC10231 présentent une croissance dans tous les puits sans échantillon ou antibiotique, ainsi que dans tous les puits du gradient de concentration en kurstakine, indiquant que la kurstakine n'a pas d'effet contre ces deux souches aux concentrations testées (de 0,4 à 200µg/mL).

## Conclusion et perspectives

Les travaux réalisés lors de cette étude sur la kurstakine ont permis d'apporter plusieurs éléments importants de compréhension du mécanisme de synthèse de ce lipopeptide par *B. thuringiensis* :

L'analyse phylogénétique de la deuxième thioestérase codée par le gène *krsD* démontre que celle-ci est proche de la famille des TeII « repair ». Le rôle de cette thioestérase est donc la réparation des domaines de thiolation de la kurstakine synthétase ayant chargé un mauvais co-facteur, entraînant alors la production d'enzymes non-fonctionnelles, ne permettant pas d'obtenir une production efficace du lipopeptide. Ainsi, si le gène *krsD* est absent ou tronqué chez une souche de *B. thuringiensis*, cela peut se traduire par un niveau de production particulièrement faible de kurstakine. A titre de comparaison, la délétion du gène codant pour la TeII « repair » associée à la surfactine synthétase chez *B. subtilis* entraîne une diminution de la production de la surfactine de 84%. La présence et l'intégrité du gène *krsD* sont donc essentielles pour assurer un taux satisfaisant de production de kurstakine chez les souches de *B. thuringiensis*.

Le maintien de la production de kurstakine par des souches chez lesquelles le gène *spo0A*, régulateur principal de la sporulation chez Bt a été inactivé démontre que la sporulation des cellules de Bt n'est pas requise pour la production du lipopeptide. Ceci implique que des conditions de culture aboutissant à la formation de spores n'est pas nécessaire à la production de kurstakine. Cet élément est particulièrement intéressant pour la mise en place de procédés de production de kurstakine impliquant l'utilisation de cultures en milieu liquide, notamment en fermenteur, associé à une technique de séparation en continu, comme cela est notamment déjà mis en place pour la purification d'autres lipopeptides.

L'augmentation de la quantité de kurstakine mesurée dans les surnageants de culture suite à la surexpression du gène *krsE* démontre que celui-ci est impliqué dans le système

d'excrétion de la molécule. A ce jour, la kurstakine n'avait jamais été dosée dans les surnageants de culture liquide de *B. thuringiensis*, et ce lipopeptide était considéré dans la littérature comme étant associé aux spores de *B. thuringiensis*. La possibilité de contrôler l'excrétion du lipopeptide par l'induction de l'expression du gène *krsE* possède un intérêt biotechnologique important pour la mise en place de procédés de productions efficaces et simples à mettre en œuvre. Ce type de procédés permettraient de purifier la molécule directement à partir du surnageant, sans avoir à l'extraire au préalable des culots cellulaires qui auront été séparés de celui-ci.

Pour les différentes souches testées, la suppression de la production de cristaux toxiques par les souches de *Bacillus thuringiensis* a permis la multiplication de la production de kurstakine par un facteur 3 à 5, indiquant un effet positif sur cette production de lipopeptide. L'une des raisons possibles de l'augmentation de la production du lipopeptide dans ces conditions pourrait être la présence d'une compétition pour l'utilisation des précurseurs composant les cristaux protéiques toxiques et le lipopeptide.

L'étude de l'augmentation de la stabilité de l'ARN messenger de la kurstakine synthétase par l'ajout d'une séquence STAB-SD en amont du gène *krsA* n'ayant abouti qu'à une légère augmentation de la production de kurstakine, il semblerait que la stabilité de cet ARNm ne soit pas un facteur limitant de la production de kurstakine. A ce jour, peu de travaux ont été réalisés sur la stabilité des ARNm de synthétases non-ribosomiques, ou sur la stabilité de ces synthétases elles-mêmes.

De plus, des avancées ont pu être réalisées au sein du laboratoire sur la purification du lipopeptide, et des éléments ont pu être apportés sur l'activité antifongique de la kurstakine.

L'étude des facteurs limitant la production de la kurstakine chez *Bacillus thuringiensis* réalisée lors de ces travaux est résumée dans le tableau

**Tableau 6 : Modifications apportées à la souche *Bacillus thuringiensis* et effet de ces modifications sur la production de kurstakine**

<b>Modification</b>	<b>Effet</b>
<b>Délétion du gène de régulation de la sporulation <i>spo0A</i></b>	La production de kurstakine est toujours effective
<b>Suppression du plasmide portant les gènes <i>cry</i>.</b>	Augmentation de la production par un facteur 3 à 5
<b>Introduction d'un stabilisateur d'ARNm en amont du locus <i>krs</i></b>	Augmentation de la production de 20%
<b>Induction de la transcription du gène <i>krsE</i></b>	Augmentation de la quantité de kurstakine mesurée dans le surnageant de culture en milieu liquide.

Ces éléments permettent alors d'envisager la mise en place d'un procédé de production et de purification de kurstakine en milieu liquide, en tirant profit des informations obtenues lors de ces travaux. L'utilisation d'une souche dont le plasmide portant les gènes *cry* aurait été supprimé, et chez laquelle l'activité du gène *krsE* pourrait être induite permettra alors de mettre en place une production en milieu liquide, et d'induire l'excrétion du lipopeptide par Bt au moment où sa production est maximale. La kurstakine pourrait alors être simplement purifiée par FPLC à partir du surnageant de culture.

Les tests d'activité antifongiques ont permis de démontrer que la kurstakine possédait une activité antagoniste contre plusieurs souches phytopathogènes. Cette information permet notamment d'envisager une nouvelle utilisation de la kurstakine dans le domaine de la phytoprotection. En effet, l'utilisation de molécules antifongiques sur des cultures peut permettre d'éviter la propagation de moisissures nuisibles des cultures. De plus, l'utilisation de molécules d'origine naturelle permet de répondre aux attentes actuelles du marché en termes de biodégradabilité des molécules utilisées, et de leur impact sur l'environnement, notamment face aux molécules issues de la synthèse chimique.

Si les travaux réalisés lors de cette étude ont permis d'apporter plusieurs éléments de compréhension sur le mécanisme de production de la kurstakine, plusieurs points restent encore à analyser. La dynamique de production de kurstakine par *B. thuringiensis* semble répondre à une logique particulière. Si la sporulation n'est pas requise pour que cette production soit active, les taux les plus importants de productions ont été observés bien après

le début de la phase stationnaire en milieu liquide, et 4 à 5 jours après le début de la culture en milieu solide, et lors de la phase stationnaire dans le cas de culture en milieu liquide. Dans le cas de cultures en milieu solide, l'augmentation du taux de kurstakine mesurée de 3 à 4 jours après le début de la croissance semble suggérer que le métabolisme des cellules de Bt est encore actif pour la production de métabolites secondaires dont la kurstakine.

L'analyse par spectrométrie de masse de kurstakine dans le surnageant de cultures de cellules de *B. thuringiensis* dans du milieu minimum M9 par Bumpus *et al.* en 2009, avait permis de détecter la coexistence de deux types de kurstakines, linéaire et cyclique. Des formes linéaires de kurstakines ont également pu être observées lors de la production de kurstakine en fermenteur décrite dans ce manuscrit. De plus, la cinétique en milieu liquide réalisée dans le cadre de ces travaux démontre une diminution de la quantité du lipopeptide mesurée dans le surnageant de culture après 24 h de production, et ce, dans les deux milieux testés. Il est donc possible que la kurstakine soit métabolisée par *B. thuringiensis* après sa synthèse, menant à la diminution drastique de la quantité du lipopeptide présent dans le surnageant, entraînant même l'impossibilité de mesurer cette concentration dans le cas du milieu LB. En raison de sa structure, la kurstakine ne représente sûrement pas une source de nutriments facilement métabolisable, et devrait donc être dégradée auparavant. Si l'augmentation de l'activité protéasique par Bt lors de la sporulation a été démontrée (Wang *et al.*, 2013), les enzymes servant à dégrader les protéines ne peuvent pas forcément attaquer facilement le cycle de la kurstakine. La linéarisation de celle-ci, aboutissant donc à l'observation de kurstakine linéaire à des temps avancés de culture dans un milieu minimum, pourrait ainsi représenter la première étape d'une série de dégradations du lipopeptide pour son recyclage par les cellules de Bt lors de leur sporulation, ou lors de la baisse de la quantité et de la disponibilité de nutriments disponibles dans le milieu de culture. Afin de déterminer la validité de cette hypothèse, un suivi complet de la production de kurstakine et l'évaluation des formes du lipopeptide détectées à partir d'une culture de *B. thuringiensis* sur un temps long sans renouvellement du milieu de culture permettrait de déterminer si les kurstakines linéaires sont directement produites par Bt, ou issues de la dégradation de kurstakines cycliques. De plus, le suivi de l'activité des gènes potentiellement impliqués dans la linéarisation ou la dégradation du lipopeptide permettrait de compléter cette étude.

## 2. Etude du mécanisme de production des PPCC

### 2.1 Introduction

Comme présenté dans l'introduction bibliographique, les PPCC regroupent plusieurs molécules présentant une forte similarité structurale. La cyclochlorotine et l'hydroxycyclochlorotine, produits par *Penicillium islandicum*, sont des pentapeptides cycliques chlorés hépatotoxiques. Les astines, connues comme étant le produit de la plante *Aster tataricus*, possèdent quant à eux de nombreuses propriétés d'intérêt, dont des activités antitumorales ou antitussives.

Toutes ces molécules sont produites en très faibles quantités, la CCT et l'HCCT nécessitant plusieurs litres de culture de *P. islandicum* pour l'obtention de quelques milligrammes de produit, et les astines nécessitant la réalisation d'extraction à partir de plusieurs kilogrammes de racines pour obtenir également quelques milligrammes.

Les gènes responsables de la production de ces produits n'étant pas connus, l'optimisation de la production de ces molécules est rendue particulièrement difficile. Le but des travaux présentés dans ce manuscrit est d'apporter de nouveaux éléments sur la connaissance du mécanisme de production de ces molécules.

Cette partie des travaux a été réalisée dans le cadre d'un projet européen ERA-IB (European Research Area for Industrial Biotechnology) intitulé Multi-Enzyme Systems Involved in Astin Biosynthesis (MESIAB), mené en collaboration avec les équipes de l'Université Technique de Dresde (Allemagne), de l'Université de Tübingen (Allemagne), de l'Université de Wageningen (Pays-Bas) et de l'Université d'Helsinki (Finlande).

Les travaux réalisés dans le cadre de cette thèse portent particulièrement sur l'identification des gènes de production des PPCC, et l'étude spécifique des domaines enzymatiques NRPS impliqués dans cette production.

## 2.2 Etude du mécanisme de production des PPCC

### 2.2.1 : Confirmation de la production de peptides de *Penicillium islandicum*

Afin de confirmer la production de cyclochlorotine et d'hydroxycyclochlorotine par *P. islandicum*, une analyse d'un surnageant d'une culture en milieu Czapek de 21 jours a été réalisée (Figure 68).

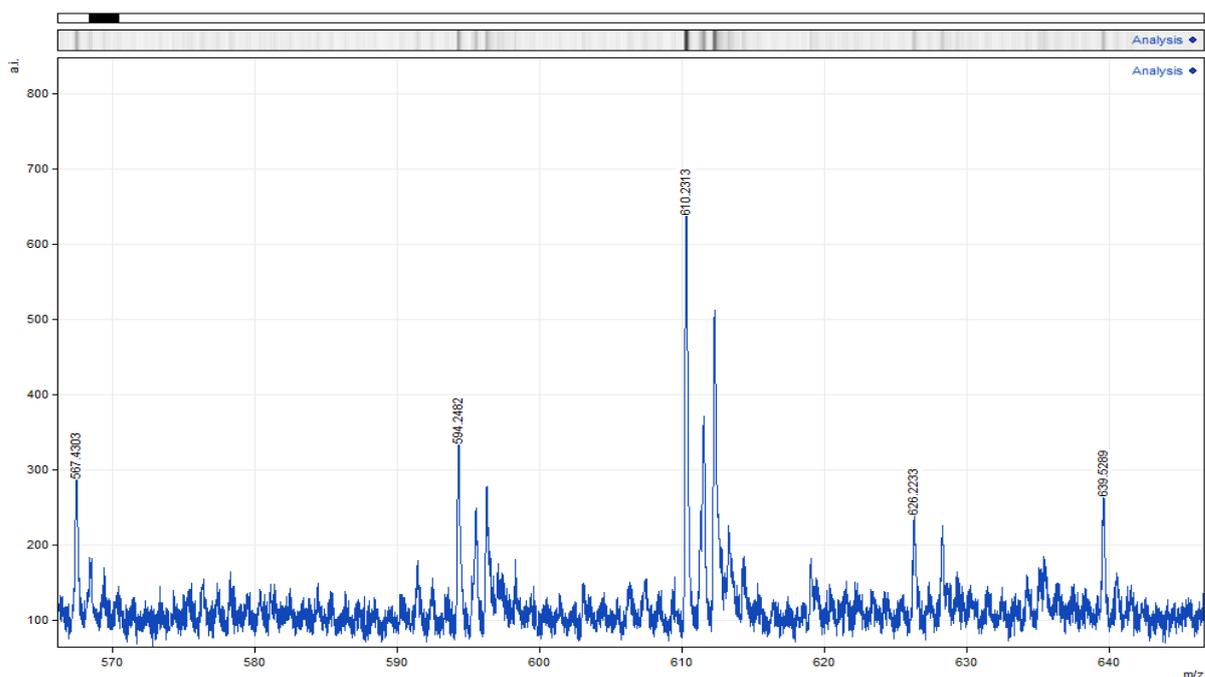


Figure 68 : Analyse par spectrométrie de masse MALDI-TOF de surnageant de culture de *P. islandicum* cultivé en milieu Czapek pendant 21 jours à 25°C.

Cette analyse permet de détecter les adduits sodiques des deux molécules étudiées, la cyclochlorotine (m/z 594,24) et l'hydroxycyclochlorotine (m/z 610,23). Les deux composés présentent un profil de massif isotopique spécifique des molécules di-chlorées.

### 2.2.2 Analyse bioinformatique du génome de *P. islandicum*

Afin de déterminer quels gènes sont responsables de la production de la cyclochlorotine et de l'hydroxycyclochlorotine chez *Penicillium islandicum*, le workflow de bioinformatique présenté dans la section Matériels et Méthodes de ce manuscrit a été employé.

### Détection des clusters dédiés à la production de métabolites secondaires

La première étape de l'analyse bioinformatique est l'identification des clusters de gènes impliqués dans la production de métabolites secondaires chez *P. islandicum*. La recherche de tels clusters est réalisée au moyen du logiciel antiSMASH. L'analyse réalisée révèle la présence de 60 clusters de métabolites secondaires répartis dans les différentes catégories détectables par antiSMASH (Tableau 7).

**Tableau 7 : Répartition des clusters impliqués dans la production de métabolites secondaires chez *P. islandicum* dans les différentes catégories par antiSMASH**

Type de cluster	Nombre de clusters
NRPS	16
PKS	19
Hybride NRPS-PKS	8
Sidérophore	1
Terpène	3
Lantanide	1
Autres	12

Le très grand nombre de clusters détectés démontre le fort potentiel de production de métabolites secondaires par *P. islandicum*. Les enzymes modulaires sont particulièrement représentées puisque les polyketide synthases représentent le type de cluster majoritaire (19 clusters), suivi par les NRPS (16 clusters) et les hybrides NRPS-PKS (8 clusters).

En raison de leur structure, la cyclochlorotine et l'hydroxycyclochlorotine sont potentiellement produites par des NRPS, celles-ci ont donc fait l'objet d'une attention particulière lors de cette étude. *P. islandicum* possède des gènes permettant de coder 16 NRPS différentes. Une analyse de l'organisation globale en domaines et modules de ces NRPS a été réalisée au moyen d'une gamme d'outils complémentaires dédiés à l'identification des modules et des domaines enzymatiques présents au sein des synthétases.

### Etude de l'organisation en domaines enzymatiques des NRPS

L'organisation en domaines enzymatiques des NRPS joue un rôle essentiel dans le mécanisme de synthèse des peptides associés. La CCT et l'HCCT étant des pentapeptides, les NRPS contenant 5 domaines d'adénylation, et donc étant possiblement responsables de la production de ces molécules, ont été sélectionnées. *P. islandicum* possède dans son génome 4

gènes, chacun impliqué dans la synthèse d'une NRPS contenant 5 domaines A. L'organisation des domaines de ces NRPS a été réalisée au moyen de plusieurs logiciels : NRPS PKS PKS NRPS et InterPro. L'organisation en domaines initialement indiquée par antiSMASH a également été analysée.

Pour chaque cluster identifié, les différents outils permettent d'obtenir un résultat différent en fonction des particularités de l'outil utilisé. Il semble donc important d'utiliser plusieurs outils et de combiner les résultats obtenus, pour proposer une interprétation la plus précise possible. A titre d'exemple, les résultats de l'analyse de l'organisation en domaines enzymatiques de la synthétase codée par le gène 2636 réalisée par les différents outils sont comparés et présentés dans le Tableau 8.

**Tableau 8 : Organisation en domaines enzymatiques de la synthétase codée par le gène 2636, déterminée par 4 outils de bioinformatique. A :Adénylation, T :Thiolation, C : Condensation, Amp : AMP-dépendant synthetase, ACP : Acyl Carrier Protein-like, ?: Non déterminé.**

Outil	Organisation en domaines de la NRPS codée par le gène 2636
<b>NRPS PKS</b>	A-T-C-A-T-?-A-T-C-A-T-C-A-T-C
<b>PKS NRPS Analysis</b>	A-T-?-A-T-?-A-T-?-A-T-?-A-T
<b>AntiSMASH</b>	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
<b>Interpro</b>	Amp-ACP-C-Amp-ACP-C-Amp-ACP-C-Amp-ACP-C-Amp-ACP-C

Les résultats fournis par NRPS-PKS permettent d'identifier un ensemble de domaines d'adénylation, de condensation et de thiolation au sein de la NRPS. En revanche, la représentation visuelle des résultats fournie par le logiciel présente certaines des liaisons entre les différents domaines en rouge (marqué par un point d'interrogation dans le Tableau 8), indiquant que la longueur entre ces deux domaines est anormalement longue et pourrait contenir un domaine n'ayant pas été détecté par le logiciel (Figure 69).



**Figure 69 : Exemple d'une liaison anormalement longue entre deux domaines enzymatiques**

Les séquences composant ces différentes liaisons feront l'objet d'une attention particulière lors de l'utilisation des autres logiciels d'analyse.

Les résultats obtenus avec le logiciel PKS-NRPS Analysis sont différents. En effet PKS-NRPS Analysis semble ne pas pouvoir détecter certains domaines de condensation et d'épimérisation. Cette incapacité pourrait être due à l'origine eucaryote des séquences analysées.

L'analyse fournie par le logiciel PKS-NRPS Analysis présente de nombreux domaines dont la fonction est inconnue (dont les domaines de condensation reconnus par NRPS-PKS) représentés par un point d'interrogation (Figure 70).

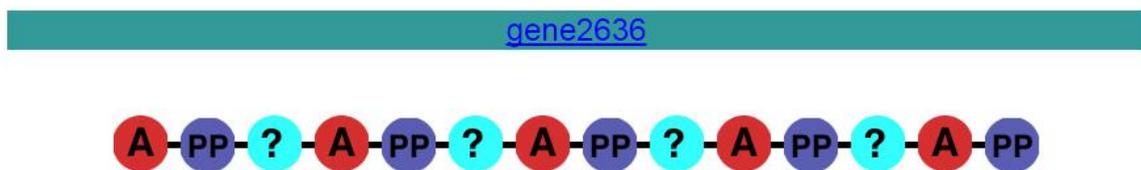


Figure 70 : Exemple d'une détection d'un domaine à la fonction inconnue lors d'une analyse par le logiciel PKS-NRPS Analysis.

En dehors de ces différences au niveau de la prédiction des domaines C, les deux logiciels fournissent une organisation similaire en domaines pour les différentes NRPS.

Les résultats de l'organisation en domaines par antiSMASH permettent d'identifier les différents domaines enzymatiques présents dans les NRPS identifiés.

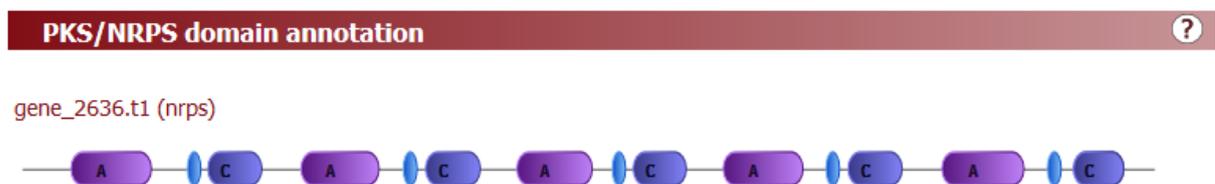


Figure 71 : Détection des domaines enzymatiques NRPS du gène 2636 par antiSMASH. Les domaines de thiolation sont représentés par les ovales bleus.

Néanmoins, des différences sont notées entre les résultats donnés par antiSMASH et les autres outils. Ainsi, pour la séquence complète de l'un des gènes codant pour une NRPS (1194), antiSMASH ne fournit qu'un domaine d'adénylation, là où PKS-NRPS Analysis détecte 14 domaines différents dont 5 domaines A à partir de la même séquence protéique.

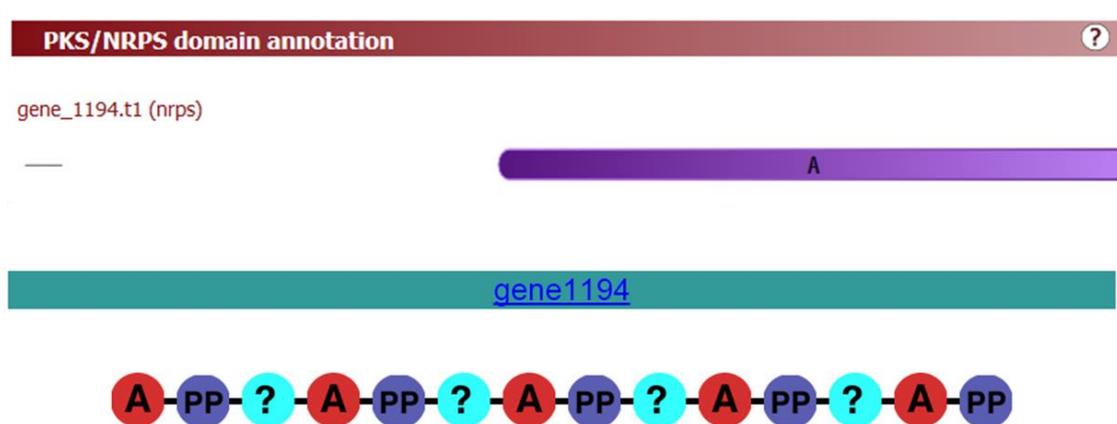


Figure 72 : Comparaison de la détection des domaines enzymatiques de la NRPS codée par le gène 1194 par antiSMASH (en haut) et PKS-NRPS Analysis (en bas).

Ce résultat indique que la détection des domaines enzymatiques par les différents outils peut parfois être très variable et qu'il est donc essentiel de combiner les résultats obtenus au moyen des différents outils.

InterProScan permet également d'identifier les différents domaines enzymatiques présents au sein des NRPS (Figure 73).

## GENE2636

Length 5,645 amino acids

### Protein family membership

None predicted.

### Domains and repeats



### Detailed signature matches



Figure 73 : Résultat d'analyse de la protéine codée par le gène 2636 à l'aide d'InterPro

Les différents domaines NRPS peuvent être identifiés au moyen des signatures associées à leur activité enzymatique. Ainsi, les domaines d'adénylation peuvent être identifiés au moyen de l'identifiant pour leur activité de synthétase AMP-dépendante (IPR000873), et parfois par

une séquence conservée à l'extrémité C-terminale de celle-ci (IPR025110). Les domaines de thiolation sont identifiés au moyen des identifiants de transporteur d'acyle (IPR009081) et de domaine de liaison du cofacteur phosphopantéthéinique (IPR020806). Enfin, les domaines de condensation sont identifiés par l'identifiant directement lié au domaine de façon générale (IPR001242).

L'utilisation d'InterPro permet ainsi d'obtenir des informations sur la nature des domaines enzymatiques présents au sein d'une NRPS malgré la défaillance d'un logiciel spécifiquement dédié à l'analyse de ces domaines.

Pour chaque domaine, les résultats obtenus avec les différents logiciels sont comparés afin de conclure sur l'organisation en domaines enzymatiques des différentes NRPS.

A l'issue de l'analyse du génome de *P. islandicum* avec les différents outils de bioinformatique, 4 gènes codant pour des NRPS possédant 5 domaines d'adénylation ont été identifiés : les gènes 1194, 2636, 3025 et 8611. Leur organisation en domaines enzymatiques est indiquée dans le tableau suivant.

**Tableau 9 : Organisation en domaines enzymatiques des NRPS possédant 5 domaines d'adénylation déduite en compilant les résultats des différents logiciels d'analyse de NRPS.**

<b>Gene/cluster</b>	<b>Organisation en domaines</b>
<b>1194</b>	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
<b>2636</b>	A-T-C-A-T-C-A-T-C-A-T-C-A-T-C
<b>3025</b>	A-T-E-C-A-T-C-A-T-E-C-A-T-C-A-T-C-C-T-C-T
<b>8611</b>	A-T-C-A-T-C-C-A-C-C-A-T-C-A-T

#### *Analyse bioinformatique de l'activité des domaines d'adénylation*

Afin de déterminer si l'un de ces clusters est impliqué dans la production de la CCT et/ou de l'HCCT, une prédiction des monomères activés par les domaines A contenus dans chacune de ces synthétases est réalisée au moyen de NRPS Predictor2, en utilisant la séquence protéique issue de chacun des gènes au format FASTA, et en sélectionnant le « prédicteur » dédié aux moisissures.

**Tableau 10 : Prédictions des monomères activés par les domaines d'adénylation des 4 NRPS identifiées chez *P. islandicum* et scores associés.**

Gene/cluster	Domaine	Three clusters	Large clusters	Small clusters	Nearest Neighbor
<b>1194</b>	A1	hydrophobic-aliphatic (0,91)	gly,ala,val,leu,ile,abu,iva (0,37)	np	Ile (50%)
	A2	hydrophobic-aliphatic (0,25)	np	np	Leu (70%)
	A3	hydrophobic-aliphatic (0,62)	gly,ala,val,leu,ile,abu,iva (0,62)	val,leu,ile,abu,iva (0,82)	Gln (60%)
	A4	Hydrophylic (0,01)	np	np	Cys (40%)
	A5	hydrophobic-aliphatic (1,10)	gly,ala,val,leu,ile,abu,iva (0,20)	np	Leu (50%)
<b>2636</b>	A1	hydrophobic-aliphatic (0,06)	Np	Np	Hpg (40%)
	A2	hydrophobic-aliphatic (0,15)	Np	Np	Leu (70%)
	A3	hydrophobic-aliphatic (0,62)	gly,ala,val,leu,ile,abu,iva (0,46)	val,leu,ile,abu,iva (0,82)	Gln (60%)
	A4		Np	Np	Cys (40%)
	A5	hydrophobic-aliphatic (1,21)	Np	Np	Leu (60%)
<b>3025</b>	A1	hydrophobic-aliphatic (0,09)	Np	Np	Lys (60%)
	A2	Np	Asp,asn,glu,gln,aad (0,12)	Np	Orn (50%)
	A3	hydrophobic-aliphatic (0,49)	Phe,trp,phg,tyr,bht (0,02)	Val,leu,ile,abu,iva (0,34)	Aad (60%)
	A4	hydrophobic-aliphatic (0,91)	gly,ala,val,leu,ile,abu,iva (0,54)	gly,ala (0,14)	Gln (50%)
	A5	hydrophobic-aliphatic (0,91)	gly,ala,val,leu,ile,abu,iva (0,84)	gly,ala (0,20)	Hpg (60%)
<b>8611</b>	A1	hydrophobic-aliphatic (1,45)	gly,ala,val,leu,ile,abu,iva (1,09)	Pro (0,20)	Val (50%)
	A2	hydrophobic-aliphatic (0,11)	Ser,thr,ser,thr,dht,dhpg,dpg,hpg (0,23)	Np	Lys (60%)
	A3	hydrophobic-aliphatic (0,58)	gly,ala,val,leu,ile,abu,iva (0,26)	Val,leu,ile,abu,iva (0,07)	Leu (60%)
	A4	hydrophobic-aliphatic (1,33)	gly,ala,val,leu,ile,abu,iva (0,52)	Val,leu,ile,abu,iva (0,15)	Asp (60%)
	A5	Np	gly,ala,val,leu,ile,abu,iva (0,17)	Np	Aad (60)

Aucune des prédictions réalisées ne permet de choisir l'un des clusters parmi les 4 comme étant le plus probablement impliqué dans la production d'une des molécules étudiées. En effet, les scores associés aux prédictions sont particulièrement faibles. Cela peut notamment être dû à la faible quantité de séquences présentes dans les bases de données associées aux prédictions chez les eucaryotes du logiciel NRPS Predictor 2 (voir paragraphe 6.3 du chapitre matériel et méthodes). Néanmoins, il est important de noter que les prédictions réalisées pour les synthétases codées par les gènes 1194 et 2636. Le modèle Nearest Neighbor ne donne ainsi qu'une seule différence entre les prédictions réalisées pour ces deux NRPS, au niveau du monomère activé par le premier domaine A de chacune d'entre elles. Ce résultat pourrait suggérer une forte similitude des peptides produits par ces enzymes.

La prédiction des peptides issus de la synthèse effectuée par les NRPS détectées n'ayant pu être effectuée avec précision, une méthode différente doit être utilisée afin de sélectionner le cluster le plus probablement impliqué dans la synthèse de la CCT et de l'HCCT.

#### *Etude du rôle joué par les domaines de condensation terminaux des NRPS de P. islandicum*

Comme indiqué dans la synthèse bibliographique, la cyclisation des peptides d'origine non-ribosomiques est effectuée chez les moisissures par un domaine de condensation terminal (C<sub>t</sub>). Les synthétases codées par les gènes 1194 et 2636 possèdent chacune à leur extrémité un domaine de condensation. Ces domaines ont été analysés de façon approfondie afin de déterminer si ceux-ci étaient bien impliqués dans la cyclisation des peptides potentiellement produits.

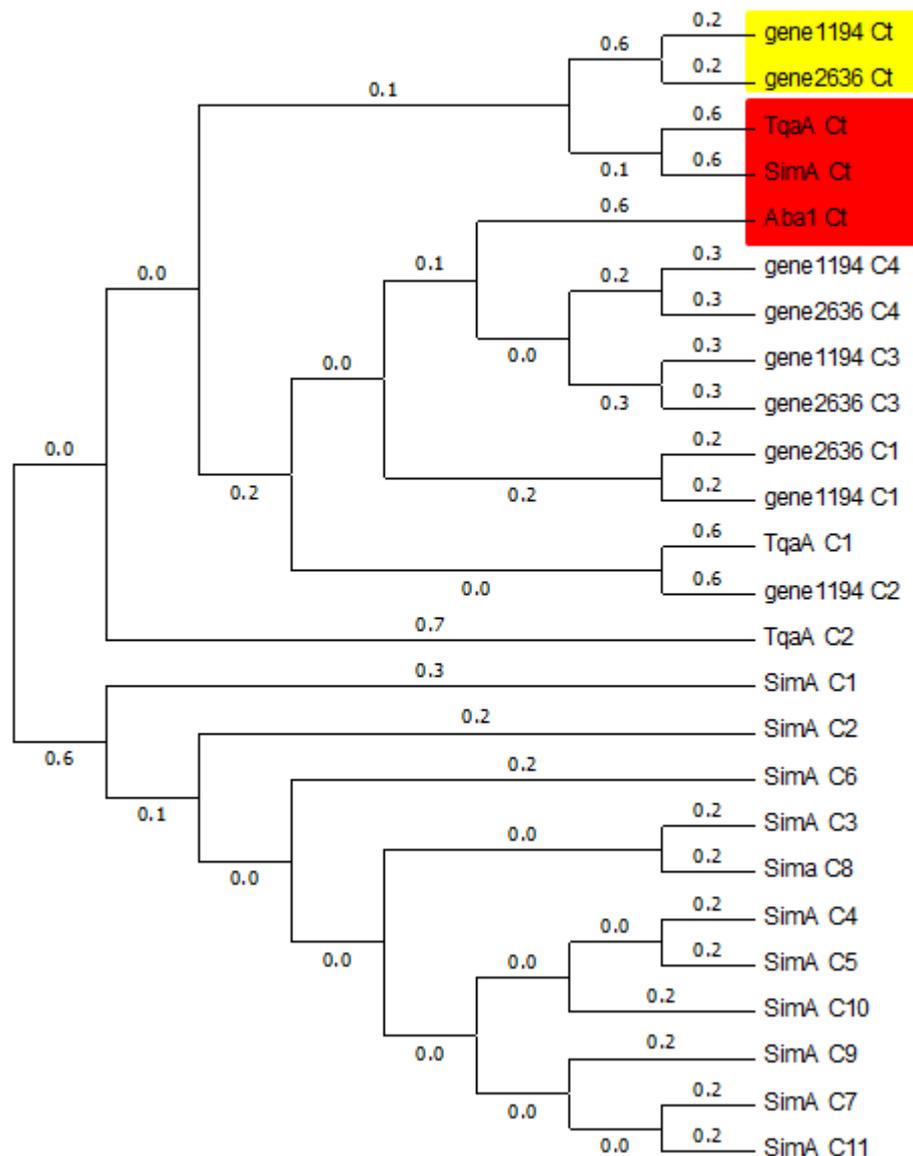
Une analyse réalisée au moyen du logiciel NaPDos indique que ces deux domaines de condensation appartiennent au sous-type des domaines dual C/E, possédant une activité de domaine C et de domaine E (Tableau 11). Ce résultat est plutôt surprenant puisqu'un domaine dual C/E ne se trouve jamais à l'extrémité d'une synthétase, car il ne peut pas condenser un monomère activé par un domaine d'adénylation devant se trouver après celui-ci.

**Tableau 11** : Résultat de l'analyse des domaines de condensation terminaux C<sub>t</sub> des deux synthétases, réalisée au moyen du logiciel NaPDos.

Query id	Database match id	percent identity	align length	e-value	pathway product	domain class
gene1194(5155)	cyclo1_C12_dual	23	354	6e-22	cyclosporin	dual
gene2636(5243)	cyclo1_C12_dual	24	358	2e-27	cyclosporin	dual

Ce résultat est dû au fait qu'à ce jour, les domaines Ct ne font pas partie des sous-types de domaines C pouvant être identifiés par NaPDoS, les plus ressemblant sont alors les dual C/E.

Comme indiqué dans la synthèse bibliographique, des travaux effectués au laboratoire ont démontré que les domaines de condensation appartenant à différents sous-types se distinguent au niveau des 150 derniers acides aminés (séquence down-seq). Afin de déterminer le rôle exact de ces deux domaines C<sub>t</sub>, leur séquence protéique a donc été alignée au moyen du logiciel MUSCLE avec les séquences de domaines C de différents sous types et l'alignement a été utilisé pour réaliser un arbre phylogénique au moyen de MEGA (Figure 74).



**Figure 74 : Arbre phylogénétique de domaines C et Ct de NRPS suspectés d’être responsables de la production de peptides cycliques chez des moisissures. Les domaines Ct issus des synthétases identifiées chez *P. islandicum* sont surlignés en jaune. Les domaines Ct issus de synthétases fongiques impliquées dans la production de peptides cycliques sont surlignés en rouge.**

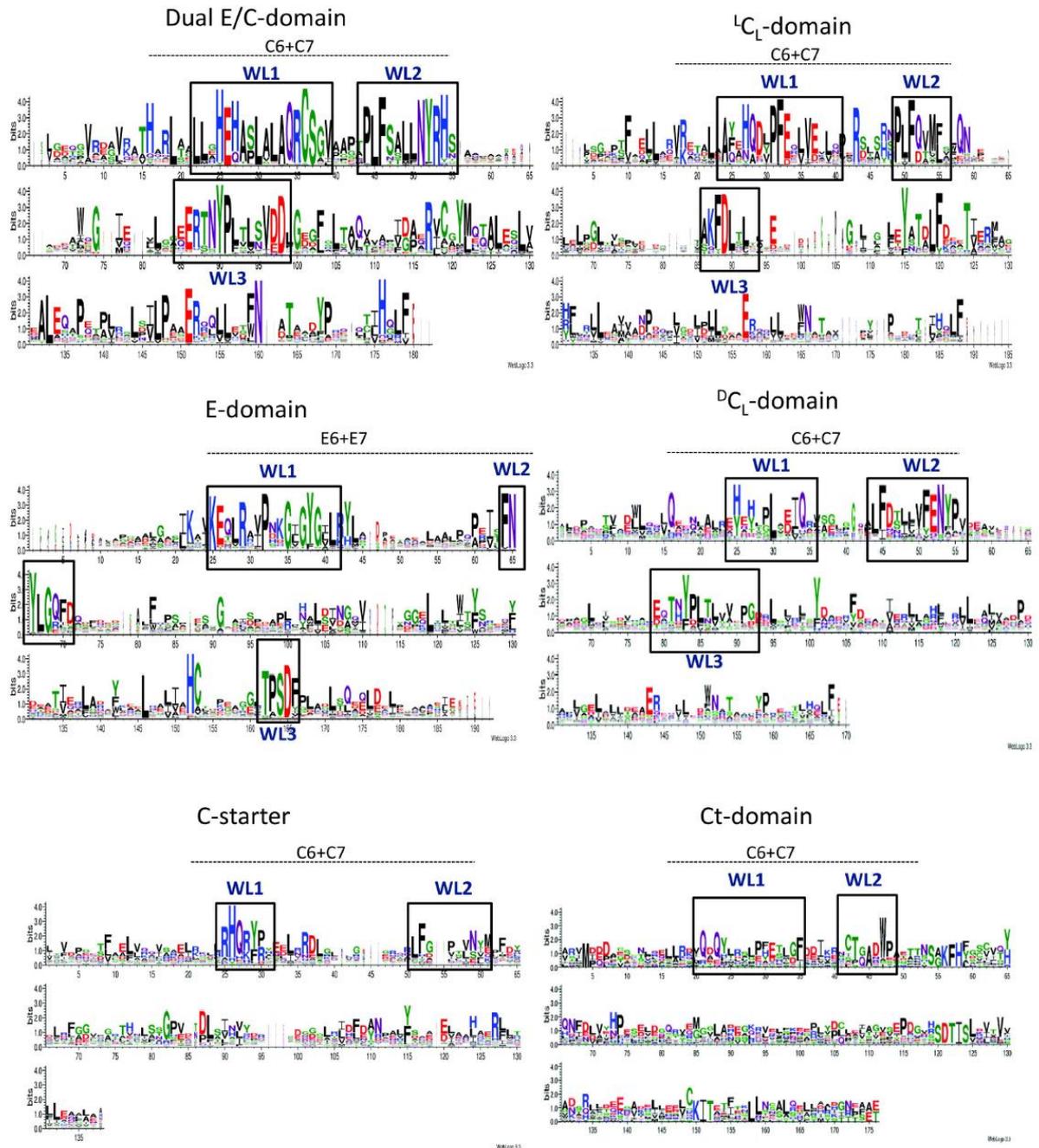
Les domaines C situés à l’extrémité des synthétases identifiées chez *P. islandicum* montrent une forte proximité avec les domaines Ct impliqués dans la cyclisation de la tryptoquialanine chez *Penicillium aethiopicum* (TqaA) et de la cyclosporine A produite par *Tolypocladium inflatum*. Ce résultat indique que les domaines C situés à l’extrémité des synthétases codées par les gènes 1194 et 2636 chez *P. islandicum* sont potentiellement impliqués dans la cyclisation des peptides produits.

### *Analyse des sous-types de domaines C*

Afin de confirmer le rôle des C terminaux présents chez *P. islandicum*, des signatures correspondant à un nouveau sous-type ont été recherchées. Ces travaux font partie d'une étude de plus grande ampleur réalisée en collaboration avec le LIFL et le LORIA pour la mise en place du workflow de bioinformatique FLORINE pour l'optimisation de l'utilisation d'outils de bioinformatique pour la prédiction des peptides issus de la voie non-ribosomique (Caradec *et al.*, 2014).

Pour rechercher des signatures pouvant être utilisées pour la reconnaissance des domaines C terminaux impliqués dans la cyclisation des peptides non-ribosomiques, une collecte de domaines C des différents sous-types a été réalisée. Un premier lot de séquences a été extrait des données de l'analyse réalisée sur les sous-types de domaines C de Rausch *et al.*, 2007. Cette étude n'ayant pas pris en compte les séquences de NPRS lorsque celles-ci provenaient de génomes partiellement séquencés, 153 séquences ont été ajoutées au jeu de données utilisé initialement dans cette étude, réparties entre les différents sous-types de domaines C (9 Cstarter, 60 <sup>L</sup>C<sub>L</sub>, 19 <sup>D</sup>C<sub>L</sub>, 31 E et 34 dual C/E.). De plus, une dizaine de domaines Ct d'origine fongique ont également été intégrés.

Des alignements de séquences down-seq de domaines C de différents sous types ont été réalisés au moyen du logiciel MUSCLE afin de déterminer des signatures permettant l'identification de ces différents domaines. Les alignements construits ont été enregistrés au format FASTA et utilisés pour la construction de Weblogos, permettant une représentation visuelle claire des positions conservées au sein des séquences analysées.



**Figure 75 :** Weblogo des différents sous-types de domaines C. Les positions des signatures C6 et C7 sont indiquées par les lignes pointillées. Les WebLogo (WL1, WL2, WL3) permettant l'identification des différents domaines sont entourés par un cadre noir.

Les WebLogo obtenus permettent d'identifier pour chaque sous type de domaine C 2 ou 3 signatures, correspondant à des régions particulièrement conservées au niveau de la séquence protéique, notées WL1, WL2 et WL3. Les signatures WL1 et WL2 correspondent bien aux régions conservées précédemment identifiées par Rausch *et al.* pour les domaines  ${}^L C_L$ ,  ${}^D C_L$ , Cstarter et Dual C/E. Par contre, la signature WL3 identifiée pour les domaines  ${}^L C_L$ ,  ${}^D C_L$ ,

Dual C/E et E est nouvelle. L'identification de ces nouvelles signatures permet d'apporter un élément nouveau pour l'identification automatique des sous-types de domaines C par les logiciels de bioinformatique dédiés.

De plus, des signatures WL1 et WL2 ont pu être identifiées pour la première fois pour les domaines Ct, qui représentent un nouveau sous-type de domaines C. Ces signatures sont différentes des signatures WL1 et WL2 identifiées pour les autres sous-types de domaine C, indiquant que les Ct possèderaient leurs propres signatures. Néanmoins, ces signatures restent relativement peu fiables comparées aux signatures des autres sous-types de domaines C, en raison de la faible quantité de séquences NRPS d'origine eucaryote impliquées dans la production de peptides cycliques référencées dans les bases de données.

La partie down-seq des Ct présents dans les synthétases chez *P. islandicum* ne contient aucune des signatures spécifiques des sous-types LCL, DLC, DualC/E, E et Cstarter, indiquant que les domaines Ct sont bien distincts des autres sous-types. Les signatures des Ct identifiés lors de cette étude n'ont également pas été retrouvées dans ces séquences. Néanmoins, la faible fiabilité de ces signatures dues au faible nombre de séquences ayant servi à leur identification ne permet pas de certifier qu'ils n'appartiennent pas à la catégorie des Ct impliqués dans la cyclisation des peptides non-ribosomiques.

#### *Bilan de l'étude bioinformatique du génome de P. islandicum*

Les différentes analyses bioinformatiques réalisées à partir du génome de *P. islandicum* ont permis d'identifier 16 clusters NRPS. Parmi ceux-ci, 4 possèdent 5 domaines d'adénylation, ce qui correspond à ce qui est recherché puisque les peptides étudiés sont constitués de 5 monomères. Les outils de prédiction pour la détermination des monomères activés par les domaines A des différentes synthétases identifiées n'ont pas permis d'identifier de synthétase potentiellement impliquée dans la synthèse de la CCT et/ou de l'HCCT en raison des faibles performances de ces logiciels pour l'analyse de données provenant d'organismes eucaryotes. L'étude de l'organisation des domaines enzymatiques de ces synthétases ainsi que l'analyse de l'activité des domaines C de celles-ci ont permis de déterminer que 2 NRPS parmi les 4 identifiées sont plus probablement responsables de la formation de peptides cycliques à 5 acides aminés. De plus, ces deux synthétases présentent un domaine de condensation terminal Ct proche des domaines Ct impliqués dans la cyclisation de peptides d'origine non-ribosomique produits par des moisissures. Ces éléments permettent de considérer ces deux synthétases comme les meilleurs candidats pour la

production de la CCT et de l'HCCT. Ces deux synthétases ont alors été retenues pour des travaux de biologie moléculaire pour l'étude de l'activité de leurs domaines A.

Les travaux de bioinformatique réalisés pour l'identification de ces synthétases ont également permis de mettre en relief les limitations des outils dédiés à l'analyse des NRPS et de leurs domaines enzymatiques lorsque celles-ci sont d'origine eucaryote. Les limites de la prédiction des monomères activés par les domaines A sont directement liées au faible nombre de séquences NRPS d'origine eucaryote disponibles. L'absence de prise en compte par les différents logiciels utilisés par cette étude de la présence de domaines Ct à la fin d'une synthétase d'origine eucaryote, entraînant la cyclisation du peptide produit pose problème quant à l'efficacité et la précision des prédictions réalisées. En effet, l'absence de la prédiction de la cyclisation du peptide produit par une telle synthétase entraîne une erreur dans la structure prédite, mais également dans sa masse et peut ainsi entraîner des difficultés pour son analyse et sa détection, notamment dans le cadre de la spectrométrie de masse.

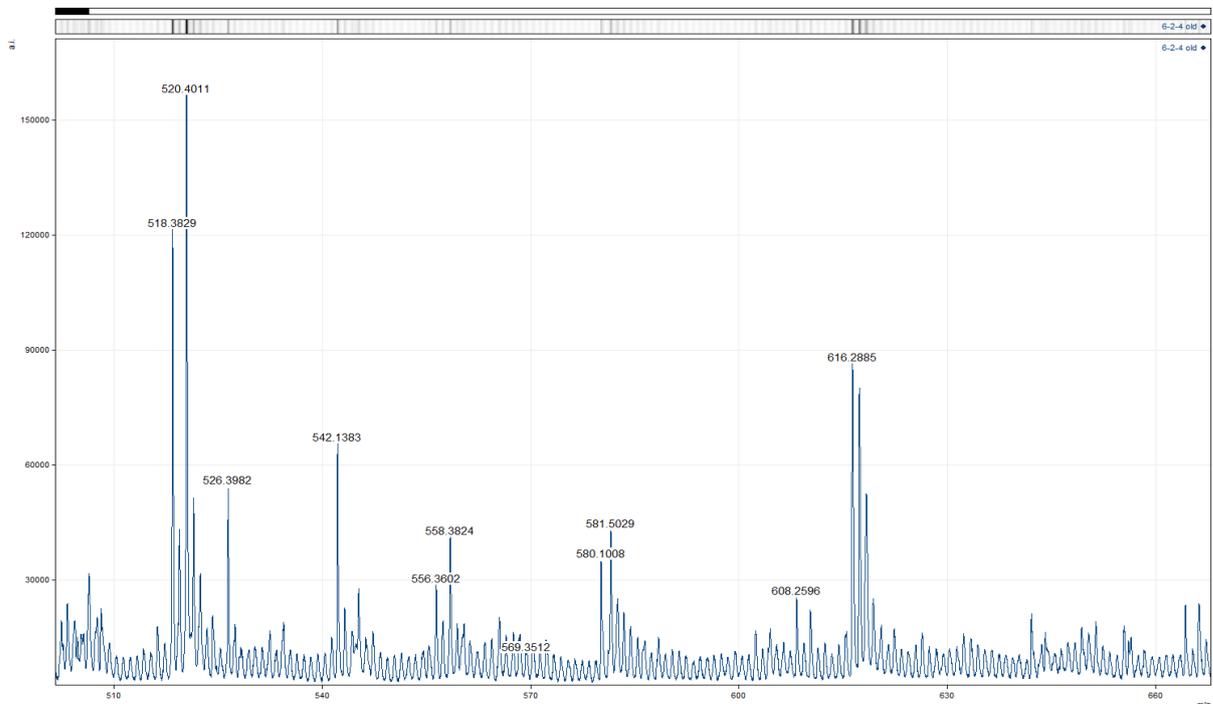
Les travaux réalisés dans le cadre de cette étude ont permis d'apporter de nouvelles données pour l'identification et la caractérisation des domaines de condensation terminaux impliqués dans la synthèse de peptides cycliques chez les champignons. Ces travaux ont permis de démontrer que les domaines Ct ne possédaient pas les signatures identifiées chez les autres sous-types de domaines C, et possédaient leurs propres signatures. Néanmoins, en raison du faible nombre de séquences NRPS responsables de la production de peptides cycliques chez les moisissures présents dans les bases de données, le nombre de séquences de domaines Ct utilisés pour l'identification de ces signatures est faible, menant à une identification de signatures encore trop peu fiable pour l'identification des domaines Ct. Néanmoins, les signatures obtenues pourront être améliorées au fur et à mesure que de nouvelles NRPS d'origine fongique impliquées dans la production de peptides cycliques seront identifiées, conduisant à l'identification plus efficace des domaines et des sous-types.

### 2.2.3 Identification d'une moisissure productrice d'astines.

Lors des travaux présentés dans ce manuscrit, une moisissure, endophyte de la plante *Aster tataricus* a été isolée par l'un des partenaires du projet MESIAB. La moisissure isolée a été identifiée comme étant un ascomycète, comme *P. islandicum*, et est proche de l'espèce *Cyanodermella viridula* (95%). La production de peptides chlorés par cette souche a été testée par spectrométrie de masse MALDI-TOF sur des surnageants de culture réalisés en milieu

Czapek en 5 semaines à 20°C sans agitation, et sur du mycélium récupéré sur milieu gélosé Czapek après 5 semaines d'incubation à 20°C.

Différentes formes d'astines ont pu être identifiées par spectrométrie de masse à partir des échantillons de mycélium (Figure 76).



**Figure 76 : Analyse par spectrométrie de masse d'un surnageant de culture de la moisissure isolée chez *P. islandicum* cultivée en milieu Czapek pendant 5 semaines à 20°C.**

Les astines B (m/z 608,26), D (m/z 558,38) et F (m/z 556,36) sont potentiellement identifiées à partir du mycélium de la moisissure. Néanmoins, le signal des pics est extrêmement faible. Aucune forme d'astine n'a pu être détectée à partir de surnageants de culture liquide.

La faiblesse des signaux détectés semble indiquer que les molécules sont produites en très faibles quantités. De plus, la croissance de la souche est particulièrement lente, une semaine de culture étant nécessaire pour l'obtention d'une fine couche de mycélium sur le dessus du milieu de culture liquide, ou sur un milieu gélosé.

Il est également important de noter que les différents partenaires du projet ont pu quant à eux identifier aux moyens d'autres techniques d'analyses des formes différentes d'astines, principalement de l'astine C.

#### 2.2.4 Analyse bioinformatique du génome de l'endophyte d'*A. tataricus*

Afin d'étudier le mécanisme de production des astines détectées chez la souche isolée à partir d'*A. tataricus*, une analyse bioinformatique similaire à celle réalisée pour le génome de *P. islandicum* a été effectuée.

Cette analyse a permis d'identifier un gène codant pour une synthétase NRPS à 5 domaines d'adénylation. Comme dans le cas des NRPS codées par les gènes 1194 et 2636 chez *P. islandicum*, la réalisation d'une analyse par le logiciel antiSMASH ne permet pas de prédire la structure du peptide issu de cette synthétase.

L'organisation en domaines enzymatique de celle-ci est similaire à celle des synthétases identifiées chez *P. islandicum*, et un domaine de condensation terminal est également présent dans cette enzyme, suggérant son implication dans la production d'un pentapeptide cyclique (Figure 77).

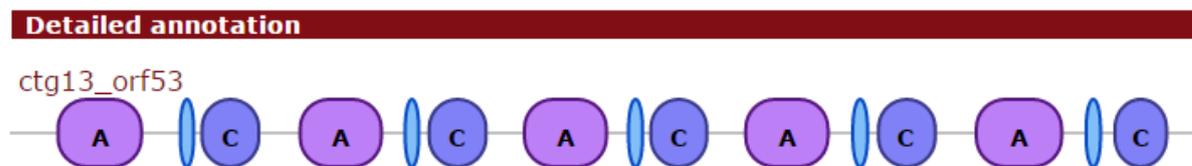


Figure 77 : Organisation en domaine de la synthétase possédant 5 domaines d'adénylation détectée dans le génome de l'endophyte d'*A. tataricus* par analyse bioinformatique.

L'alignement des gènes codant pour les 3 synthétases démontre une forte similarité des séquences.

Néanmoins, comme pour les synthétases codées par les gènes 1194 et 2636 chez *P. islandicum*, l'analyse bioinformatique réalisée sur la synthétase identifiée chez l'endophyte ne permet pas de prédire le peptide produit par cette NRPS.

#### 2.2.5 Analyse de l'activité des domaines A des synthétases de *P. islandicum*.

L'analyse bioinformatique n'ayant pas permis d'identifier avec certitude le ou les gènes codant les synthétases responsables des produits de *P. islandicum*, deux autres stratégies ont été utilisées. Elles visent à définir la spécificité des domaines d'adénylation présents dans ces synthétases.

- La première est une méthode d'analyse *in vitro*, consistant à cloner les domaines d'adénylation chez différents hôtes pour une expression hétérologue de ces domaines enzymatiques permettant l'analyse de leur activité.
- La seconde est une méthode d'analyse *in vivo*, consistant à modifier l'apport en acides aminés composant les peptides étudiés, ou précurseur des acides aminés composant ces peptides.

#### *Travaux de biologie moléculaire*

Les travaux de biologie moléculaire réalisés dans le cadre de l'étude des domaines A des NRPS identifiés chez *P. islandicum* n'ont pas permis de cloner les gènes codant ces domaines. En effet, après extraction de l'ADN de *P. islandicum* et digestion enzymatique de celui-ci, aucun clone n'a pu être obtenu par la méthode de gap-repair cloning réalisé chez *Saccharomyces cerevisiae* et décrite dans le paragraphe 4.4 du chapitre Matériel et Méthodes.

La méthode de clonage par restriction enzymatique (voir paragraphe 4.3 du chapitre Matériel et Méthode) utilisé pour la production de domaines A hétérologues chez *P. pastoris* n'a également pas permis d'obtenir de domaines A solubles et fonctionnels. Si plusieurs clones intermédiaires ont été obtenus chez *E.coli*, ceux-ci se sont révélés être de faux positifs après tentative d'amplification par PCR de l'amplicon inséré dans le vecteur plasmidique.

De plus, une tentative similaire d'expression hétérologue de ces domaines réalisée par l'un des partenaires du projet MESIAB chez 3 souches d'*E. coli* (Rosetta, BL21 et KRX) n'ont également pas permis de réaliser cette analyse, peu de clonages réalisés ayant permis une surexpression des protéines recombinantes, et les seules protéines obtenues n'étant pas soluble.

Enfin, une tentative de réalisation de mutant de délétion des gènes NRPS étudiés chez *P. islandicum* a également été réalisée sans succès à ce jour par un autre partenaire du projet.

Les effets de l'alimentation différenciée en acides aminés sur la production des métabolites secondaires produits par *P. islandicum* et de possibles dérivés de ces métabolites ont été étudiés. Pour cela, la moisissure a été mise en culture dans du milieu Czapek modifié, dans lequel un acide aminé a été ajouté en excès (voir Matériel et méthodes). Chaque culture a été réalisée pendant 21 jours à 25°C sans agitation. Le surnageant de culture est analysée à l'issue de cette culture par spectrométrie de masse MALDI-TOF.

L'analyse des différents échantillons par spectrométrie de masse permet de détecter les deux molécules CCT et HCCT. Elle permet aussi de détecter d'autres molécules présentant des masses et des profils isotopiques différents.

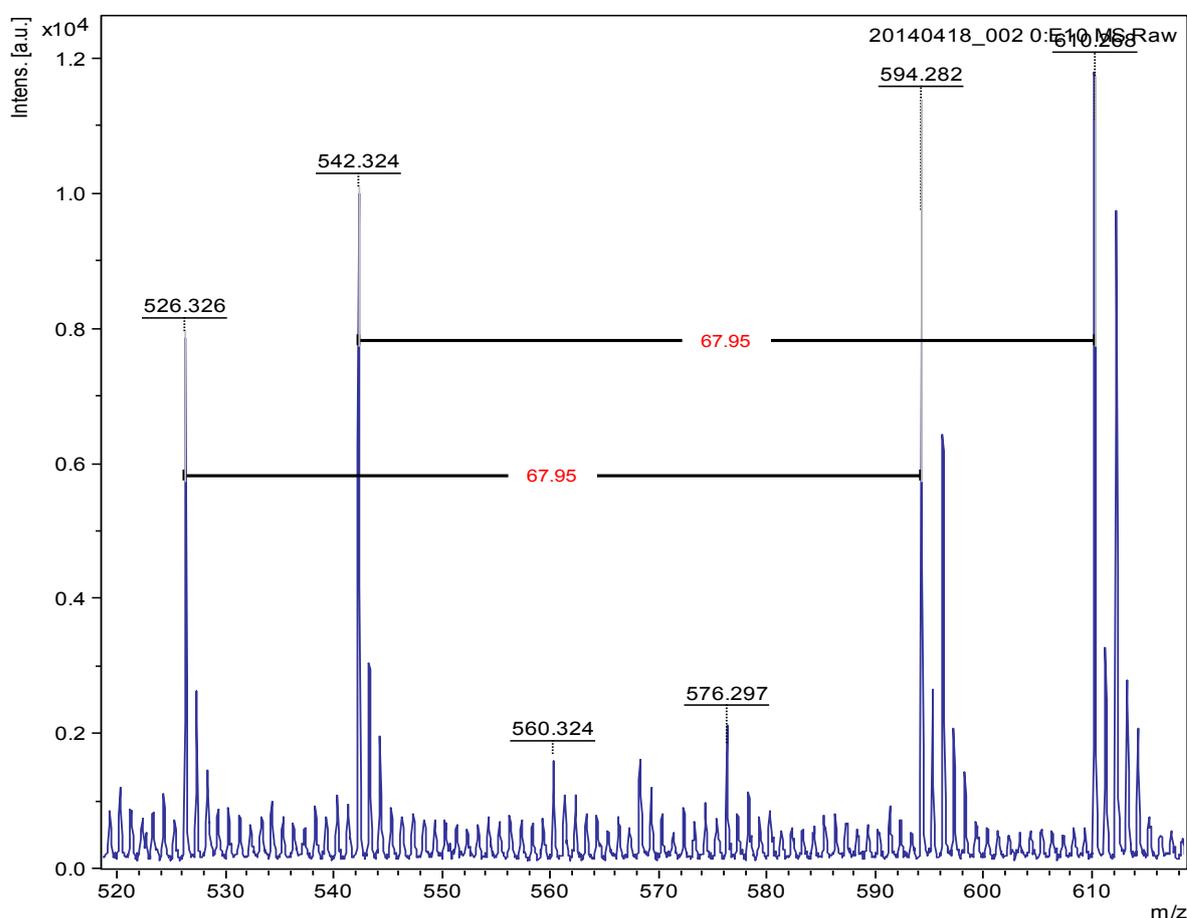


Figure 78 : Analyse par spectrométrie de masse des différents pentapeptides produits par *P. islandicum* en milieu liquide Czapek après 21 jours à 25°C.

Deux pics d'intensité importante sont détectés aux masses 526 et 542. Chacun de ces pics présente une différence de masse de 67,85 avec la CCT et l'HCCT respectivement. De plus, ces différents pics ne présentent pas de profil caractéristique des molécules chlorées.

Ces observations suggèrent que ces pics seraient ceux de dérivés non-chlorés des deux molécules étudiées. En effet, la différence de masse correspond à l'absence des deux chlores (-70 environ), remplacés chacun par un atome d'hydrogène (+2 environ).

Deux autres pics de masse  $m/z$  560 et 576 d'intensité faible peuvent également être détectés. La faible intensité du signal de ces pics rend toute tentative d'observation d'un profil du massif isotopique difficile. Néanmoins, la différence de masse présentée entre ces molécules et la CCT et l'HCCT d'une part, et les deux dérivés non-chlorés supposés d'autre part, permet d'émettre l'hypothèse de la présence de dérivés des deux molécules ne comportant qu'un atome de chlore (Figure 79).

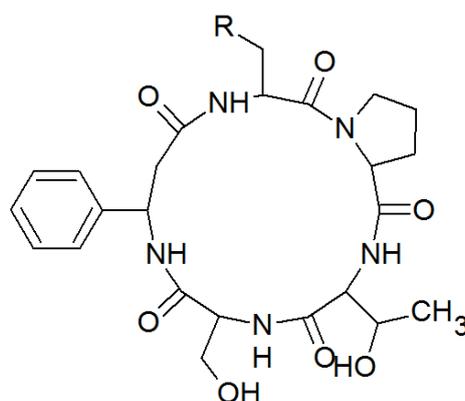


Figure 79 : Structure proposée pour les pics détectés en MALDI-TOF à  $m/z$  526 (R=H) et 542 (R=OH).

C'est à notre connaissance la première fois que ces molécules sont détectées dans le surnageant de culture de *P. islandicum*.

Dans les différentes conditions testées, les 3 groupes de molécules sont détectés. Les molécules considérées comme monochlorées ne sont toujours présente qu'en quantité extrêmement faible, amenant à la présence d'un signal extrêmement faible.

Afin d'en apprendre plus sur le lien entre les différentes molécules détectées, le ratio des différents niveaux d'halogénéation de celles-ci a été mesuré pour chaque échantillon en comparant les intensités des différents pics.

Les ratios mesurés permettront ainsi d'observer l'évolution de la proportion entre les deux molécules initialement étudiées (HCCT/CCT) et entre leur dérivé non chloré (CT/HCT).

Une autre série de mesures permettra cette fois d'observer plus attentivement l'évolution de l'halogénéation des différentes molécules en comparant les intensités de chaque molécule et de son dérivé non chloré.

En raison de la faible intensité des pics des molécules mono-chlorés supposées, les valeurs de ces intensité ne sera pas utilisé pour cette étude, l'effet du bruit de fond de l'analyse risquant alors de grandement influencer ces valeurs.

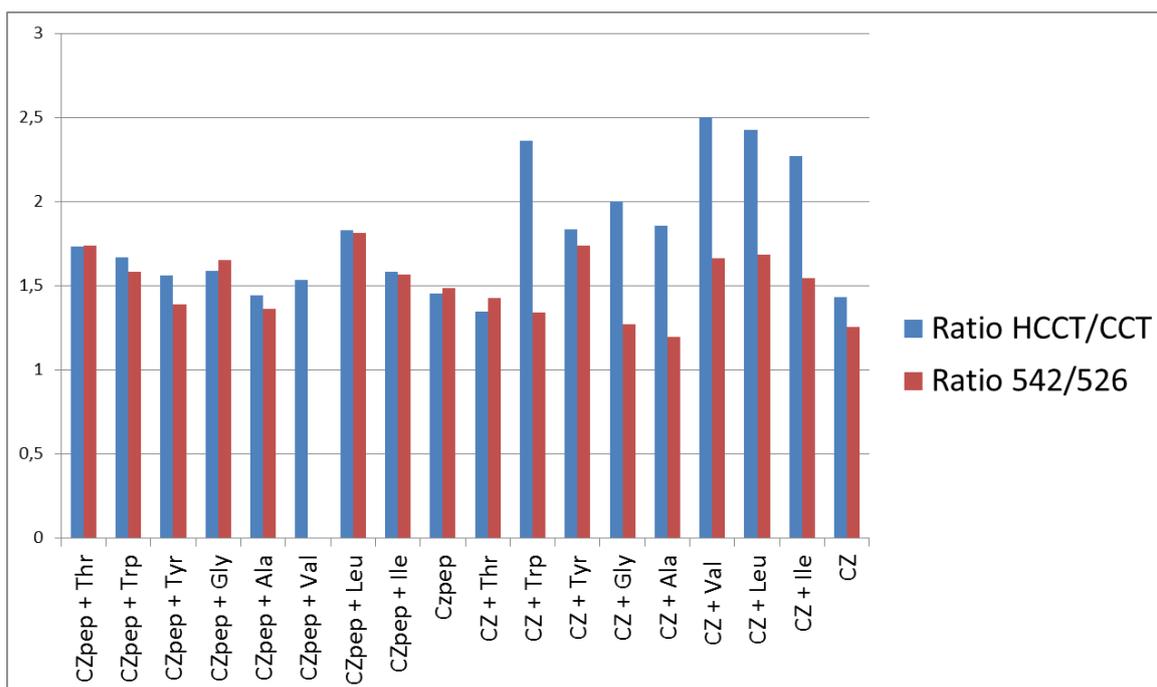
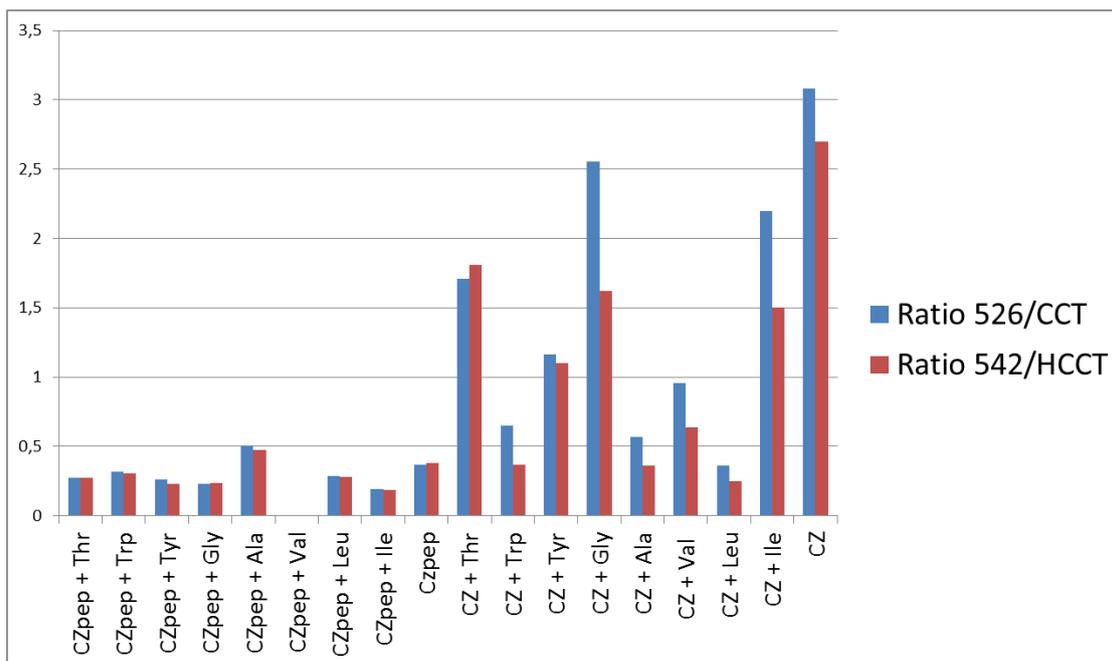


Figure 80 : Ratio de l'intensité des pics observés en spectrométrie de masse MALDI-TOF d'échantillons de surnageant de culture de *Penicillium islandicum* cultivé en milieu Czapek additionné d'un acide aminé en excès pendant 21 jours à 25°C.

Le ratio HCCT/CCT évolue peu dans les différentes conditions testées. Toutes les valeurs sont supérieures à 1, indiquant que quelles que soient les conditions testées, l'HCCT est toujours présente en quantité plus importante que la CCT. Ceci se retrouve également au niveau des molécules non chlorées, l'intensité du pic à 542 m/z étant toujours supérieur à celle du pic à 526 m/z.

Ces résultats indiquent que l'alimentation différenciée en acide aminé n'a pas entraîné de modification notable dans le ratio des molécules



**Figure 81 : Ratios entre les molécules non chlorées et les molécules chlorées. CZpep : Czapek peptone ; CZ : Czapek.**

La comparaison des ratios des molécules dichlorées et des molécules non-chlorées apporte des éléments de compréhension intéressants sur le mécanisme de production des molécules.

Dans le cas des milieux Czapek modifiés contenant de la peptone, tous les ratios 526/CCT et 542/HCCT sont inférieurs à 1, indiquant que les molécules non chlorées sont toujours présentes en quantités supérieures aux molécules chlorées. Lorsque le milieu ne contient pas de peptone, le ratio est différent. En effet, lorsque le milieu est additionné de certains acides aminés, les molécules non-chlorées sont majoritaires (exemple de l'ajout de thréonine ou de glycine pour lequel les ratios sont supérieurs à 1), tandis que d'autres acides aminés produisent l'effet inverse (exemple du tryptophane ou de l'alanine pour lesquels les ratios sont inférieurs à 1).

Il est également important de noter que pour les ratios 526/CCT et 542/HCCT semblent évoluer de la même façon en fonction du milieu. L'alimentation différentielle en acides aminés semble donc modifier l'activité d'halogénéation des peptides par *Penicillium islandicum*.

A l'issue de cette expérience, deux faits peuvent être soulignés :

- Quel que soit le milieu utilisé, les molécules portant une thréonine sont toujours présentes en quantité supérieure à celles portant un acide aminobutyrique.
- L'ajout d'acides aminés peut modifier l'activité d'halogénéation des peptides étudiés.

## Conclusion et perspectives

La recherche des gènes de production de la cyclochlorotine et de l'hydroxycyclochlorotine chez *Penicillium islandicum* a permis l'identification de deux clusters de gènes aboutissant à la production de NRPS potentiellement impliquées dans la synthèse de ces molécules. L'analyse bioinformatique réalisée pour l'identification de ces clusters de gènes a amené à soulever plusieurs éléments importants concernant l'utilisation des outils bioinformatiques dédiés à l'analyse des NRPS.

En premier lieu, certains outils ne sont pas optimaux en termes de détection et de qualification des domaines enzymatiques de ces NRPS. En effet, les domaines de condensation de ces clusters n'ont pas toujours été correctement identifiés par certains de ces outils tels NRPS-PKS Analysis ou PKS-NRPS. L'outil NaPDoS, quant à lui, n'a pas permis la caractérisation efficace des domaines Ct, impliqués dans la cyclisation des peptides non-ribosomiques chez les moisissures. La défaillance de ces outils a requis l'utilisation de méthodes différentes pour analyser les différents clusters. Ainsi, une étude phylogénétique des domaines C terminaux des clusters identifiés a permis de confirmer que ceux-ci sont effectivement proches des domaines Ct connus chez d'autres moisissures, supposant alors leur activité de cyclisation. Une étude plus poussée de ces domaines a également permis d'identifier pour la première fois des signatures spécifiques de ces domaines dans la partie « down-seq » de ceux-ci, les différenciant des autres sous-types de domaines C connus, chez qui des signatures spécifiques avaient déjà été identifiées. L'identification de ces signatures pourrait permettre à terme d'implémenter les outils de bioinformatique dédiés à l'étude de NRPS de fonctions permettant de reconnaître les domaines Ct, permettant la prise en compte de leur présence pour la prédiction de la structure du peptide synthétisé par cette synthétase.

En deuxième lieu, l'utilisation du logiciel de prédiction des monomères activés par les domaines d'adénylation NRPS Predictor 2 n'a pas permis de prédire les peptides produits par les différents clusters. Les prédictions réalisées par le logiciel sont accompagnées de scores particulièrement faibles, soulignant le manque de fiabilité de la prédiction réalisée. Ces scores faibles sont explicables par le faible nombre de séquences de NRPS d'origine eucaryote composant la base de données utilisée par le logiciel pour les prédictions.

Au cours de ces travaux, la découverte d'une moisissure endophyte d'*A. tataricus* productrice d'astines par l'un des partenaires du projet lors de cette étude a permis pour la première fois d'attribuer la production d'astines à un organisme différent d'*A. tataricus*. L'étude du génome de cette moisissure a également permis d'identifier un gène NRPS codant pour une synthétase possédant 5 domaines d'adénylation, mais dont le produit n'a pu être déterminé par bioinformatique, pour les mêmes raisons que pour *Penicillium islandicum*.

Afin de confirmer l'implication des synthétases identifiées chez *P. islandicum* et l'endophyte d'*A. tataricus*, des travaux ont été réalisés pour étudier en détail l'activité des domaines d'adénylation de ces synthétases. D'une part, des travaux de biologie moléculaire dont le but était l'expression hétérologue de ces domaines pour l'étude *in vitro* de leur activité enzymatique, et d'autre part, des expériences d'alimentation différenciée en acides aminés pour l'étude *in vivo* de l'activité des domaines A.

Les travaux de biologie moléculaire réalisée dans le cadre de ces travaux et par nos partenaires n'ont pas permis de valider l'implication d'un des gènes codant pour une NRPS identifiés dans le génome de *P. islandicum* comme responsable de la production de CCT et/ou d'HCCT.

Les expériences d'alimentation différenciée ont permis de démontrer que l'ajout de différents acides aminés ne permettait pas d'obtenir de modification de la séquence peptidique des molécules produites, mais qu'une modification du niveau d'halogénéation des peptides pouvait être observée.

A l'issue de ces travaux, une hypothèse sur le mécanisme de production de la CCT et de l'HCCT chez *P. islandicum* peut être émise :

Les deux NRPS codées par les gènes 1194 et 2636 identifiées dans le génome de la souche étant particulièrement proche et étant tous deux impliqués dans la synthèse d'un

pentapeptide cyclique, chacune pourrait être responsable de la synthèse d'un des peptides étudiés. Les différentes modifications de milieux n'ayant pas permis de modifier de façon significative le ratio entre les deux molécules ou leurs dérivés, les deux gènes codant pour ces synthétases pourraient être soumises au même système de régulation. En revanche, l'ajout de différents acides aminés semble modifier le degré d'halogénéation des deux molécules de façon homogène, suggérant qu'une même halogénase est impliquée dans la modification des deux peptides.

# **CHAPITRE 4**

## **Conclusion**

Les travaux présentés dans ce manuscrit avaient comme but d'apporter des éléments de compréhension sur le mécanisme de synthèse de deux familles de molécules, les kurstakines et les astines. Ces deux familles de molécules laissent entrevoir un potentiel extrêmement intéressant pour leur valorisation industrielle, notamment en raison de leur activité biologique connue, telle l'activité antitumorale des astines, ou en cours d'identification, comme l'activité antifongique des kurstakines.

Une meilleure compréhension des mécanismes de production de ces peptides devaient ainsi permettre d'améliorer les taux de production de ces peptides, jusqu'ici produite en faible quantité, et nécessitant des procédés de production laborieux et difficile à mettre en place, et d'obtenir ainsi des quantités suffisantes de peptides pour confirmer leur potentiel.

Ces travaux représentent aussi le moyen de mettre en valeur les différences entre les mécanismes de synthèse non-ribosomique entre les organismes procaryotes et eucaryotes.

### **La kurstakine : une molécule soumise à une régulation complexe, de la transcription à l'excrétion**

Les travaux sur la kurstakine ont tout d'abord permis de sélectionner une souche sauvage présentant le meilleur taux de production de kurstakine et encore peu étudiée dans le cadre de cette production, la souche *Bacillus thuringiensis* HD-73. Cette souche a été utilisée pour réaliser une série de modifications génétiques permettant de mieux comprendre le mécanisme de production de la kurstakine et l'interaction de ce mécanisme avec d'autres voies métaboliques de la souche.

La détection du lipopeptide chez une souche non sporulante, mutée au niveau du régulateur principal de la sporulation *spo0A* nous a permis de démontrer que la sporulation des cellules de *B. thuringiensis* n'est pas nécessaire à la production du lipopeptide.

Par une analyse bioinformatique, le rôle potentiel de la deuxième thioestérase codée par le gène *krsD* a été défini comme étant celui d'une TeII « repair », impliquée dans la réparation des domaines de thiolation portant un mauvais co-substrat, aboutissant à la formation d'une synthétase non fonctionnelle. Le gène *krsD* a donc été identifié comme non-essentiel à la production de kurstakine, mais important pour l'obtention d'un taux élevé de production du lipopeptide, comme cela est le cas pour les TeII « repair » associées à la surfactine synthétase et à l'arthrofactine synthétase.

La délétion des gènes de production de cristaux toxiques chez *B. thuringiensis* a permis d'augmenter de manière significative le taux de production de kurstakine. Cette augmentation pourrait être due à l'augmentation des nutriments et précurseurs nécessaires à la production de lipopeptide, lorsque ceux-ci ne sont pas utilisés pour la production de toxines.

Le rôle du gène *krsE* a été défini au moyen d'une souche chez laquelle le promoteur du gène *krsE* est inductible. L'augmentation de la quantité de kurstakine dans le surnageant de culture de cette souche lors de l'induction de la transcription de ce gène indique que la protéine KrsE est impliquée dans l'excrétion du lipopeptide dans le milieu de culture.

Les travaux réalisés ont donc permis d'apporter des éléments de compréhension du mécanisme de production de la kurstakine par *B. thuringiensis*, ceci permettra à terme la mise en place d'un premier procédé de production de kurstakine en milieu liquide au moyen d'une souche modifiée.

Les principaux éléments de compréhension du mécanisme de production de la kurstakine apportés au moyen de ces travaux concernent les différentes étapes de ce mécanisme :

- La transcription des gènes du locus *krs*, et la synthèse de l'ARNm
- La modification post-traductionnelle, avec le passage de la kurstakine synthétase de la forme apo-enzyme à la forme holo-enzyme.
- La disponibilité des précurseurs pour la production du lipopeptide.
- L'excrétion de la molécule, via la protéine KrsE.
- La dégradation rapide de la molécule.

Ces différents éléments sont synthétisés dans la figure suivante :

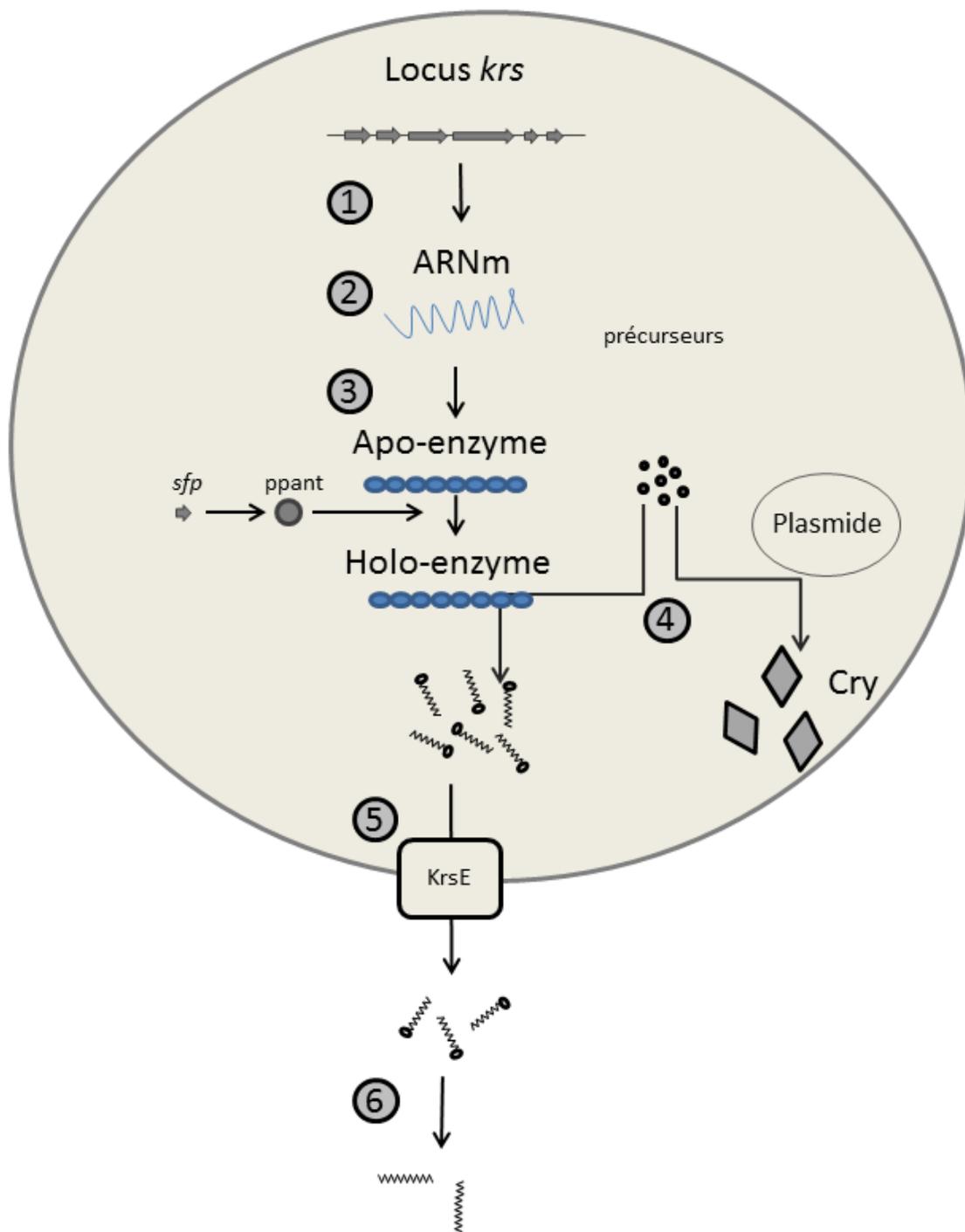


Figure 82 : Schéma récapitulatif du mécanisme de production de la kurstakine par *Bacillus thuringiensis* et des informations récoltées au cours de cette étude. 1 : Les gènes *krsABC* du locus *krs* sont transcrits en ARN messenger. 2 : L'augmentation de la stabilité de l'ARNm permet l'augmentation de la production de kurstakine. 3 : L'ARNm est traduit en une synthétase non fonctionnelle, sous forme d'apoenzyme, rendue fonctionnelle par l'action de la ppant codée par le gène *sfp*. 4 : Les précurseurs sont utilisés pour la production de lipopeptides et/ou de toxines protéiques sous forme de cristaux. 5 : L'excrétion de la kurstakine est effectuée par la protéine KrsE. 6 : Les lipopeptides sont dégradés dans le milieu de culture.

Plusieurs éléments restent encore à étudier pour la compréhension du mécanisme de production de ce lipopeptide. Parmi ceux-ci la possibilité d'une mauvaise transcription des gènes *sfp* et *krsD*, pouvant aboutir à la production d'un faible nombre de synthétases fonctionnelles. Une étude sur le niveau de transcription des différents gènes du locus *krs* permettrait de déterminer si ces deux gènes jouent un rôle de facteur limitant dans la production de kurstakine.

De la même façon, la dégradation du lipopeptide entraînant la diminution de la quantité de cette molécule dans le milieu de culture observée lors des cinétiques de production réalisées permettent d'envisager l'action d'une enzyme permettant cette dégradation. L'identification de cette enzyme et la délétion du gène dont celle-ci serait issue permettrait également le maintien de la quantité de lipopeptide.

Enfin, l'analyse de l'activité biologique a permis de confirmer le caractère antifongique de ce lipopeptide contre des souches phytopathogènes, confirmant l'intérêt écologique et économique de cette molécule.

### **Travaux sur les PPCC : identification de clusters de gènes potentiels et travaux sur les domaines d'adénylation.**

Les travaux sur les PPCC ont permis d'identifier deux gènes de synthétase non-ribosomique potentiellement responsable de la production de la cyclochlorotine et/ou de l'hydroxycyclochlorotine chez *P. islandicum*, et également un gène NRPS potentiellement responsable de la production de l'astine chez l'endophyte isolé à partir des racines d'*A. tataricus*.

Les limites des outils de bioinformatique n'ayant pas permis de prédire avec précision les peptides issus de ces synthétases, et donc de confirmer l'implication de ces NRPS dans cette production de peptides, une stratégie d'analyse de l'activité des domaines d'adénylation a été mise en place. Cette analyse a été divisée en deux parties, une partie *in vitro* consistant en un clonage et une expression hétérologue des domaines A des NRPS pour l'analyse de leur activité, et une partie *in vivo* consistant en une série d'expérience d'alimentation différenciée en acides aminés pour l'étude des peptides résultant de la synthèse non-ribosomiques obtenus dans ces conditions.

Les techniques de biologie moléculaire mises en place au laboratoire n'ont pas permis d'obtenir des domaines A solubles nécessaires pour l'étude de leur activité, et ce type d'expérience mené par nos partenaires, ainsi que des expériences de mutagenèse dirigée sur les deux gènes NRPS identifiés chez *P. islandicum* effectuée par notre partenaire de Tübingen au sein de ce projet n'ont pas permis non plus de confirmer l'implication des gènes 1196 et 2636 dans la production de la cyclochlorotine et de l'hydroxycyclochlorotine.

Les expériences d'alimentation différenciée en acides aminés n'ont pas permis d'obtenir des variants des molécules étudiés chez lesquels la séquence peptidique différencierait de celle des peptides déjà connus. Néanmoins, ces expériences ont permis de démontrer que le ratio entre la CCT et l'HCCT variait peu quelles que soient les conditions testées, mais également d'identifier des dérivés non-chlorés et mono-chlorés des deux molécules, jusqu'alors jamais détectés dans des surnageants de culture de *P. islandicum*.

## **Bilan**

Les travaux présentés dans ce manuscrit ont permis d'apporter de nouveaux éléments et de nouvelles pistes pour la compréhension du mécanisme de synthèse des deux familles de molécules étudiées, la kurstakine et les PPCC.

Avant le début de ces travaux, le niveau de connaissances sur les deux sujets était très inégal. Là où les gènes de productions et une partie du mécanisme de régulation de ceux-ci étaient connus pour la production de kurstakine chez *B. thuringiensis*, aucun élément n'avait été apporté à ce jour sur les gènes de production de la CCT, de l'HCCT ou de l'astine.

Les travaux, réalisés sur un modèle procaryote d'une part, et eucaryote d'autre part, ont également permis de mettre en relief les grandes différences entre les travaux sur ces deux types de modèle :

D'un point de vue microbiologique, les temps de culture et de production des molécules étudiés sont extrêmement différents. Là où 24 à 48H sont suffisantes pour réaliser l'identification d'un peptide produit par une bactérie au temps de croissance rapide, les travaux sur les moisissures ont demandés de 3 à 5 semaines de croissance.

D'un point de vue bioinformatique, les outils dédiés à l'analyse des NRPS se sont souvent révélés bien plus efficaces pour les procaryotes que les eucaryotes. En effet, que cela concerne l'analyse des domaines enzymatiques NRPS ou la prédiction des monomères activés

par des domaines d'adénylation, les outils ont été moins efficaces pour l'analyse du génome de *P. islandicum* que pour *B. thuringiensis*. Cela est notamment dû au fait que l'apprentissage de ces outils est essentiellement réalisé au moyen de séquences provenant d'organisme procaryote, souvent plus étudiés, et plus simples à manipuler que les organisme eucaryotes.

Ainsi, parmi les peptides référencés dans la base de données NORINE, 63% sont d'origine bactérienne. Pourtant, l'analyse du génome de *P. islandicum* et de l'endophyte d'*A. tataricus* dévoile de très forts potentiels de production NRPS, que l'on retrouve également chez d'autres souches séquencées, certains organismes, comme *Aspergillus fumigatus* ou *Aspergillus nidulans*, pouvant comprendre plus d'une dizaine de clusters de gènes NRPS dans leur génome (Stack et al., 2007 ; Von Döhren, 2009).

Les difficultés d'analyse des génomes eucaryotes pour l'identification des peptides issus de la voie non-ribosomique soulignent l'importance du développement et de l'amélioration des outils de bioinformatique dédiés à ces analyses, comme l'identification de signatures spécifiques aux domaines Ct réalisée dans le cadre de ces travaux.

## Références Bibliographiques

- Abderrahmani, A., Tapi, A., Nateche, F., Chollet, M., Leclère, V., Wathelet, B., Hacene, H., and Jacques, P. (2011). Bioinformatics and molecular approaches to detect NRPS genes involved in the biosynthesis of kurstakin from *Bacillus thuringiensis*. *Appl. Microbiol. Biotechnol.* 92, 571–581.
- Agaisse, H., Gominet, M., Okstad, O.A., Kolstø, A.B., and Lereclus, D. (1999). PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol. Microbiol.* 32, 1043–1053.
- Ansari, M.Z., Yadav, G., Gokhale, R.S., and Mohanty, D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res* 32, W405–W413.
- Balibar, C.J., Vaillancourt, F.H., and Walsh, C.T. (2005). Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* 12, 1189–1200.
- Béchet, M., Caradec, T., Hussein, W., Abderrahmani, A., Chollet, M., Leclère, V., Dubois, T., Lereclus, D., Pupin, M., and Jacques, P. (2012). Structure, biosynthesis, and properties of kurstakins, nonribosomal lipopeptides from *Bacillus* spp. *Appl Microbiol Biotechnol* 95, 593–600.
- Bučević-Popović, V., Sprung, M., Soldo, B., and Pavela-Vrančić, M. (2012). The A9 core sequence from NRPS adenylation domain is relevant for thioester formation. *Chembiochem* 13, 1913–1920.
- Bumpus, S.B., Evans, B.S., Thomas, P.M., Ntai, I., and Kelleher, N.L. (2009). A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* 27, 951–956.
- Caboche, S., Pupin, M., Leclère, V., Fontaine, A., Jacques, P., and Kucherov, G. (2008). NORINE: a database of nonribosomal peptides. *Nucleic Acids Res.* 36, D326–331.
- Caradec, T., Pupin, M., Vanvlassenbroeck, A., Devignes, M.-D., Smail-Tabbone, M., Jacques, P., and Leclère, V. (2014). Prediction of Monomer Isomery in Florine: A Workflow Dedicated to Nonribosomal Peptide Discovery. *PLoS ONE* 9, e85667.
- Chen, M.L., and Tsen, H.Y. (2002). Discrimination of *Bacillus cereus* and *Bacillus thuringiensis* with 16S rRNA and *gyrB* gene based PCR primers and sequencing of their annealing sites. *J. Appl. Microbiol.* 92, 912–919.
- Chino, A., Watanabe, K., and Moriya, H. (2010). Plasmid construction using recombination activity in the fission yeast *Schizosaccharomyces pombe*. *PLoS ONE* 5, e9652.
- Von Döhren, H. (2009). A survey of nonribosomal peptide synthetase (NRPS) genes in *Aspergillus nidulans*. *Fungal Genetics and Biology* 46, S45–S52.

- Du, C., and Nickerson, K.W. (1996). *Bacillus thuringiensis* HD-73 Spores Have Surface-Localized Cry1Ac Toxin: Physiological and Pathogenic Consequences. *Appl. Environ. Microbiol.* *62*, 3722–3726.
- Dubois, T., Faegri, K., Perchat, S., Lemy, C., Buisson, C., Nielsen-LeRoux, C., Gohar, M., Jacques, P., Ramarao, N., Kolstø, A.-B., et al. (2012). Necrotrophism Is a Quorum-Sensing-Regulated Lifestyle in *Bacillus thuringiensis*. *PLoS Pathog* *8*, e1002629.
- Finking, R., and Marahiel, M.A. (2004). Biosynthesis of nonribosomal peptides1. *Annu. Rev. Microbiol.* *58*, 453–488.
- Gao, X., Haynes, S.W., Ames, B.D., Wang, P., Vien, L.P., Walsh, C.T., and Tang, Y. (2012). Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nat. Chem. Biol.* *8*, 823–830.
- Gevers, W., Kleinkauf, H., and Lipmann, F. (1968). The activation of amino acids for biosynthesis of gramicidin S. *Proc Natl Acad Sci U S A* *60*, 269–276.
- Ghosh, A.C., Manmade, A., Kobbe, B., Townsend, J.M., and Demain, A.L. (1978). Production of luteoskyrin and isolation of a new metabolite, pibasterol, from *Penicillium islandicum* Sopp. *Appl. Environ. Microbiol.* *35*, 563–566.
- Gurib-Fakim, A. (2006). Medicinal plants: traditions of yesterday and drugs of tomorrow. *Mol. Aspects Med.* *27*, 1–93.
- Hathout, Y., Ho, Y.P., Ryzhov, V., Demirev, P., and Fenselau, C. (2000). Kurstakins: a new class of lipopeptides isolated from *Bacillus thuringiensis*. *J. Nat. Prod.* *63*, 1492–1496.
- Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., and Kolstø, A.B. (2000). *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*--one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* *66*, 2627–2630.
- Ibrahim, M.A., Griko, N., Junker, M., and Bulla, L.A. (2010). *Bacillus thuringiensis*: a genomics and proteomics perspective. *Bioeng Bugs* *1*, 31–50.
- Ishii, K., and Ueno, Y. (1973). Production of the hepatotoxic chlorine-containing peptide by *Penicillium islandicum* Sopp. *Appl Microbiol* *26*, 359–363.
- Ishikawa, I., Ueno, Y., and Tsunoda, H. (1970). Chemical determination of the chlorine-containing peptide, a hepatotoxic mycotoxin of *Penicillium islandicum* Sopp. *J. Biochem.* *67*, 753–758.
- Itokawa, H., Morita, H., Nagashima, S., and Takeya, K. (1994). Cyclic Peptides from Higher Plants. Part 8. Three Novel Cyclic Pentapeptides, Astins F, G and H from *Aster tataricus*. *HETEROCYCLES* *38*, 2247.
- Jacques, P. (2011). Surfactin and Other Lipopeptides from *Bacillus* spp. In *Biosurfactants*, G. Soberón-Chávez, ed. (Springer Berlin Heidelberg), pp. 57–91.
- Jegorov, A., Hajduch, M., Sulc, M., and Havlicek, V. (2006). Nonribosomal cyclic peptides: specific markers of fungal infections. *J Mass Spectrom* *41*, 563–576.

- Jiang, Z., Vasil, A.I., Hale, J.D., Hancock, R.E.W., Vasil, M.L., and Hodges, R.S. (2008). Effects of net charge and the number of positively charged residues on the biological activity of amphipathic alpha-helical cationic antimicrobial peptides. *Biopolymers* 90, 369–383.
- Joung, K.-B., and Côté, J.-C. (2002). A Review of Environmental Impacts of the Microbial Insecticide *Bacillus thuringiensis*: A54-9/29 - Government of Canada Publications.
- Kleinkauf, H., and von Döhren, H. (1990). Nonribosomal biosynthesis of peptide antibiotics. *Eur. J. Biochem.* 192, 1–15.
- Koglin, A., Löhr, F., Bernhard, F., Rogov, V.V., Frueh, D.P., Strieter, E.R., Mofid, M.R., Güntert, P., Wagner, G., Walsh, C.T., et al. (2008). Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* 454, 907–911.
- Kohli, R.M., Trauger, J.W., Schwarzer, D., Marahiel, M.A., and Walsh, C.T. (2001). Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry* 40, 7099–7108.
- Kosemura, S., Ogawa, T., and Totsuka, K. (1993). Isolation and Structure of Asterin, a new halogenated cyclic penta-peptide from *Aster tataricus*. *Tetrahedron Letters* 34, 1291–1294.
- Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* 3, 923–936.
- Lee, B.-N., Kroken, S., Chou, D.Y.T., Robbertse, B., Yoder, O.C., and Turgeon, B.G. (2005). Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryotic Cell* 4, 545–555.
- Liu, X., Cao, P., Zhang, C., Xu, X., and Zhang, M. (2012). Screening and analyzing potential hepatotoxic compounds in the ethanol extract of *Asteris Radix* by HPLC/DAD/ESI-MSn technique. *Journal of Pharmaceutical and Biomedical Analysis* 67–68, 51–62.
- Madonna, A.J., Voorhees, K.J., Taranenko, N.I., Laiko, V.V., and Doroshenko, V.M. (2003). Detection of Cyclic Lipopeptide Biomarkers from *Bacillus* Species Using Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* 75, 1628–1637.
- Maget-Dana, R., Thimon, L., Peypoux, F., and Ptak, M. (1992). Surfactin/iturin A interactions may explain the synergistic effect of surfactin on the biological properties of iturin A. *Biochimie* 74, 1047–1051.
- Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* 97, 2651–2674.
- Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 39, W339–W346.

- Mizumoto, S., Hirai, M., and Shoda, M. (2007). Enhanced iturin A production by *Bacillus subtilis* and its effect on suppression of the plant pathogen *Rhizoctonia solani*. *Appl. Microbiol. Biotechnol.* *75*, 1267–1274.
- Mizutani, K., Hirasawa, Y., Sugita-Konishi, Y., Mochizuki, N., and Morita, H. (2008). Structural and conformational analysis of hydroxycyclochlorotine and cyclochlorotine, chlorinated cyclic peptides from *Penicillium islandicum*. *J. Nat. Prod.* *71*, 1297–1300.
- Morita, H., Nagashima, S., Takeya, K., and Itokawa, H. (1993a). Astins A and B, antitumor cyclic pentapeptides from *Aster tataricus*. *Chem. Pharm. Bull.* *41*, 992–993.
- Morita, H., Nagashima, S., Shiota, O., Takeya, K., and Itokawa, H. (1993b). Two Novel Monochlorinated Cyclic Pentapeptides, Astins D and E from *Aster tataricus*. *Chemistry Letters* 1877–1880.
- Morita, H., Nagashima, S., Takeya, K., and Itokawa, H. (1994). A Novel Cyclic Pentapeptide with a  $\beta$ -Hydroxy- $\gamma$ -chloroproline from *Aster tataricus*. *Chemistry Letters* *23*, 2009–2010.
- Morita, H., Nagashima, S., Takeya, K., and Itokawa, H. (1995). Structure of a new peptide, astin J, from *Aster tataricus*. *Chem. Pharm. Bull.* *43*, 271–273.
- Peypoux, F., Bonmatin, J.M., and Wallach, J. (1999). Recent trends in the biochemistry of surfactin. *Appl. Microbiol. Biotechnol.* *51*, 553–563.
- Price, N.P.J., Rooney, A.P., Swezey, J.L., Perry, E., and Cohan, F.M. (2007). Mass spectrometric analysis of lipopeptides from *Bacillus* strains isolated from diverse geographical locations. *FEMS Microbiol. Lett.* *271*, 83–89.
- Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* *33*, 5799–5808.
- Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D.H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology* *7*, 78.
- Razafindralambo, H., Popineau, Y., Deleu, M., Hbid, C., Jacques, P., Thonart, P., and Paquot, M. (1997). Surface-Active Properties of Surfactin/Iturin A Mixtures Produced by *Bacillus subtilis*. *Langmuir* *13*, 6026–6031.
- Romero, D., de Vicente, A., Rakotoaly, R.H., Dufour, S.E., Veening, J.-W., Arrebola, E., Cazorla, F.M., Kuipers, O.P., Paquot, M., and Pérez-García, A. (2007). The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. *Mol. Plant Microbe Interact.* *20*, 430–440.
- Roongsawang, N., Washio, K., and Morikawa, M. (2010). Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. *Int J Mol Sci* *12*, 141–172.

- Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011). NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* 39, W362–W367.
- Samel, S.A., Marahiel, M.A., and Essen, L.-O. (2008). How to tailor non-ribosomal peptide products—new clues about the structures and mechanisms of modifying enzymes. *Mol Biosyst* 4, 387–393.
- Schafer, P. (2011). *The Chinese Medicinal Herb Farm: A Cultivator's Guide to Small-scale Organic Herb Production* (Chelsea Green Publishing).
- Schneider, A., and Marahiel, M.A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch Microbiol* 169, 404–410.
- Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., and Dean, D.H. (1998). *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* 62, 775–806.
- Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14083–14088.
- Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. *Nat Prod Rep* 20, 275–287.
- Shen, Y., Luo, Q., Xu, H., Gong, F., Zhou, X., Sun, Y., Wu, X., Liu, W., Zeng, G., Tan, N., et al. (2011). Mitochondria-dependent apoptosis of activated T lymphocytes induced by astin C, a plant cyclopeptide, for preventing murine experimental colitis. *Biochem. Pharmacol.* 82, 260–268.
- Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 6, 493–505.
- Stack, D., Neville, C., and Doyle, S. (2007). Nonribosomal peptide synthesis in *Aspergillus fumigatus* and other fungi. *Microbiology* 153, 1297–1306.
- Strieker, M., Tanović, A., and Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics. *Current Opinion in Structural Biology* 20, 234–240.
- Tapi, A., Chollet-Imbert, M., Scherens, B., and Jacques, P. (2010). New approach for the detection of non-ribosomal peptide synthetase genes in *Bacillus* strains by polymerase chain reaction. *Appl. Microbiol. Biotechnol.* 85, 1521–1531.
- Tatsuno, T., Tsukioka, M., Sakai, Y., Suzuki, Y., and Asami, Y. (1955). [Study of the toxic principle in yellow rice]. *Pharm Bull* 3, 476–477.
- Terao, K., Ito, E., and Tatsuno, T. (1984). Liver injuries induced by cyclochlorotine isolated from *Penicillium islandicum*. *Arch. Toxicol.* 55, 39–46.
- Tomino, S., Yamada, M., Itoh, H., and Kurahashik, null (1967). Cell-free synthesis of gramicidin S. *Biochemistry* 6, 2552–2560.

- Ueno, Y. (1992). [Hepatotoxicity of cyclochlorotine--a cyclic peptide produced by *Penicillium islandicum*]. *Prikl. Biokhim. Mikrobiol.* 28, 899–906.
- Ueno, Y., and Ishikawa, I. (1969). Production of Luteoskyrin, a Hepatotoxic Pigment, by *Penicillium islandicum* Sopp. *Appl Microbiol* 18, 406–409.
- Vanittanakom, N., Loeffler, W., Koch, U., and Jung, G. (1986). Fengycin--a novel antifungal lipopeptide antibiotic produced by *Bacillus subtilis* F-29-3. *J. Antibiot.* 39, 888–901.
- Weber, T., and Marahiel, M.A. (2001). Exploring the Domain Structure of Modular Nonribosomal Peptide Synthetases. *Structure* 9, R3–R9.
- Weckermann, R., Fürbass, R., and Marahiel, M.A. (1988). Complete nucleotide sequence of the *tycA* gene coding the tyrocidine synthetase 1 from *Bacillus brevis*. *Nucleic Acids Res* 16, 11841.
- Xinrong, Y. (2003). *Encyclopedic Reference of Traditional Chinese Medicine* (Springer Science & Business Media).
- Xu, H.-M., Zeng, G.-Z., Zhou, W.-B., He, W.-J., and Tan, N.-H. (2013). Astins K–P, six new chlorinated cyclopentapeptides from *Aster tataricus*. *Tetrahedron* 69, 7964–7969.
- Yoshioka, H., Nakatsu, K., Sato, M., and Tatsuno, T. (1973). The molecular structure of cyclochlorotine, a toxic chlorine-containing cyclic pentapeptide. *Chemistry Letters* 1319–1322.

INDEX PHYTOSANITAIRE ACTA 2013.