
Université de Lille (Faculté des Sciences et Technologies)
&
Université de Liège – Gembloux Agro-Bio Tech

Thèse

Présentée par Antoine VASSAUX

Pour l'obtention du titre de :

Docteur de l'Université de Lille en Ingénierie des fonctions biologiques,

Docteur de l'Université de Liège Gembloux Agro-Bio Tech en Sciences
agronomiques et Ingénierie biologique

Mécanisme de biosynthèse et production de l'astine, un pentapeptide cyclique non-ribosomique de *Cyanodermella asteris*

Préparée à l'Institut Charles Viollette et au TERRA Teaching and Research Centre

Soutenue le 24 Septembre 2019 devant le Jury composé de :

Valérie LECLERE	(Université de Lille, France)	Directrice
Patrick FICKERS	(Université de Liège, Belgique)	Directeur
Philippe JACQUES	(Université de Liège, Belgique)	Co-Directeur
Micheline VANDENBOL	(Université de Liège, Belgique)	Présidente
Sylvie LAUTRU	(Université de Paris Saclay, France)	Rapportrice
Stéphane COCIANCICH	(CIRAD de Montpellier, France)	Rapporteur
Carole Molina JOUVE	(Université de Toulouse, France)	Examinatrice
Sébastien RIGALI	(Université de Liège, Belgique)	Examinateur

Mécanisme de biosynthèse et production de l'astine, un pentapeptide cyclique non-ribosomique de *Cyanoderrella asteris*

Antoine Vassaux



Promoteurs :

Université de Lille : Valérie LECLERE

Université de Liège Gembloux Agro-Bio Tech : Patrick FICKERS, Philippe JACQUES

Année civile : 2019

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLoux AGRO-BIO TECH

Et

UNIVERSITÉ DE LILLE (FRANCE)
FACULTÉ DES SCIENCES ET TECHNOLOGIES (ST)

**Mécanisme de biosynthèse et production
de l'astine, un pentapeptide cyclique
non-ribosomique de *Cyanodermella asteris***

Antoine Vassaux

Dissertation originale présentée en vue de l'obtention du grade de docteur de
l'Université de Liège Gembloux Agro-Bio Tech en sciences agronomiques et
ingénierie biologique, et de l'Université de Lille en Ingénierie des fonctions
biologiques

Promoteurs :

Université de Lille : Valérie LECLERE

Université de Liège Gembloux Agro-Bio Tech : Patrick FICKERS, Philippe JACQUES

Année civile : 2019

« Le succès n'est pas final, l'échec n'est pas fatal,
C'est le courage de continuer qui compte. »

Winston Churchill

Résumé

L'astine C est un peptide cyclique assemblé au cours d'un mécanisme non-ribosomique par une synthétase dite NRPS (NonRibosomal Peptide Synthetase). Ce peptide non-ribosomique possède des propriétés thérapeutiques intéressantes avec notamment des activités anti-tumorale et anti-inflammatoire. Jusqu'ici ce composé était exclusivement extrait à partir des racines d'*Aster tataricus*, une plante utilisée traditionnellement en médecine japonaise et chinoise. Récemment, une production d'astine C a été mise en évidence chez *Cyanoderrella asteris*, un champignon endophyte de cette plante. Cette découverte ouvre de nouvelles perspectives en matière de production d'astine C, qui reste néanmoins très limitée en raison du faible taux de croissance du champignon endophyte. Au cours de cette étude, deux approches ont été développées parallèlement afin d'augmenter les taux de production de l'astine C.

La première stratégie consistait à augmenter les rendements en optimisant la production homologue à partir de *C. asteris*. Dans cette optique, un système de culture a été établi afin de cultiver le champignon exclusivement sur un support en acier inoxydable. Ce mode de culture a favorisé à la fois le développement de la biomasse fongique et la production du composé d'intérêt. En vue d'optimiser ce procédé, l'impact de plusieurs paramètres de culture (modalité de préparation du support, type d'inoculum, pH de culture, et composition du milieu de culture) sur la production d'astine C a été évalué. Les paramètres de culture optimisés ont permis d'améliorer de nouveau les rendements en astine C, ce qui constitue une première étape dans le développement d'un procédé de production à l'échelle industrielle.

En parallèle, des travaux ont été menés afin de développer un système de production hétérologue d'astine C chez la levure. Cette approche n'a pu être considérée qu'après avoir identifié, par le biais d'analyses bioinformatiques, les gènes impliqués dans la voie de biosynthèse de l'astine. Une fois ces gènes identifiés, une revue de la littérature a permis, entre autres, de dresser un bilan des outils moléculaires disponibles pour le clonage des larges séquences nucléiques codant pour les NRPSs, et de sélectionner des hôtes hétérologues appropriés. Des séquences complètes ou partielles du gène codant pour l'astine synthétase ont été clonées respectivement chez *Saccharomyces cerevisiae* et *Yarrowia lipolytica*. Pour les deux levures considérées, une expression hétérologue a été constatée. Chez *S. cerevisiae*, la synthèse de la NRPS d'astine n'a pas pu être démontrée. En revanche, pour la première fois, la production d'une structure de type NRPS chez *Y. lipolytica* a pu être mise en évidence. Bien qu'aucun peptide non-ribosomique n'ait été détecté, cette étude a permis de lever une partie des verrous limitant le développement d'un mode de production hétérologue d'astine chez la levure.

Abstract

Astin C is a cyclic peptide assembled through a nonribosomal mechanism by a NonRibosomal Peptide Synthetase (NRPS). This nonribosomal peptide displays promising therapeutic properties including anti-tumor and anti-inflammatory activities. For decades, this compound was only extracted from the roots of *Aster tataricus*, a well-known plant in traditional Japanese and Chinese medicine. Recently, *Cyanodermella asteris*, a fungal endophyte of this plant, was demonstrated to be able to synthesize astin C. This discovery offers new opportunities for the production of this compound of interest. Nonetheless, the very low growth rate of this endophytic fungus is an obstacle limiting the astin C production. In this study, two distinct approaches were jointly considered to upscale the production rate of this compound.

The first strategy was related to an optimization of the homologous production from *C. asteris*. For this purpose, an innovative cultivation system has been developed enabling to grow the fungus exclusively on a stainless steel support. This cultivation condition turned out to be favorable both for the fungal biomass development and for the astin C production. In order to further optimize this process, the effects of several culture parameters (i.e. support pre-treatment procedure, inoculum type, pH, medium composition) on the astin C production rates was investigated. Under optimized conditions, astin C yields were further enhanced, constituting a first step towards the development of an astin C production process at industrial scale.

Meanwhile, a study was conducted in order to develop a heterologous expression system for astin C production in yeast. The identification, through bioinformatics analyses, of the genes involved in the astin biosynthesis was a precondition for the development of this approach. Once these genes have been identified, a literature review has enabled to compile the molecular tools applicable for the cloning of NRPS long length sequence, and to select a proper host to heterologously express them. The whole sequence or a truncated one have been transferred respectively to *Saccharomyces cerevisiae* or *Yarrowia lipolytica*. In both considered yeasts, a heterologous expression of the foreign sequences was confirmed. In *S. cerevisiae*, the synthesis of the astin NRPS could not be demonstrated. On the other hand, in *Y. lipolytica*, for the first time, the production of a NRPS-type structure was detected. Although no nonribosomal peptide was detected, this study enabled to overcome several of the hurdles limiting the development of a astin heterologous production way in yeast.

Remerciements

« Se réunir est un début ; rester ensemble est un progrès ; travailler ensemble est la réussite ». Cette citation d'Henry Ford pourrait sans doute constituer le slogan de ce projet de thèse... Le travail de recherche que je vous présente ici est le fruit d'une collaboration transfrontalière entre l'université de Lille et l'université de Liège-Gembloux ayant rendu possible la réalisation de cette thèse en cotutelle. Parmi les caractéristiques d'une cotutelle, mis à part l'augmentation du nombre de déplacements (et des déménagements), on peut citer la possibilité de multiplier les rencontres, les opportunités, et les collaborations. Ayant très largement usé et abusé de cette possibilité, j'ai un bon nombre de remerciements à adresser aux personnes que j'ai côtoyées de part et d'autre de la frontière, et même au-delà :

À commencer par l'ensemble des membres du jury qui ont accepté d'évaluer mon travail réalisé au cours de la thèse. Je remercie vivement les Professeurs Sylvie Lautru et Stéphane Cociancich pour le temps qu'ils ont consacré pour évaluer et corriger ce manuscrit. Un grand merci également à Carole Jouve, Micheline Vandebol et Sébastien Rigali d'avoir accepté d'examiner ce travail et de participer au jury d'évaluation de la thèse.

Je tiens ensuite à exprimer ma gratitude à Valérie Leclère, ma directrice de thèse à l'université de Lille, ainsi qu'à Patrick Fickers et Philippe Jacques, mes promoteurs/co-promoteurs à l'université de Liège-Gembloux. Je vous remercie chaleureusement tous les trois pour tout ce que vous m'avez apporté scientifiquement ces dernières années, mais également pour votre soutien inconditionnel et votre bonne humeur à toute épreuve tout au long du projet et ce malgré les (quelques) obstacles.

J'adresse des remerciements tout particuliers à Micheline Vandebol pour m'avoir accueilli une partie de la thèse au sein de son unité de Microbiologie et Génomique à Gembloux, pour son aide précieuse dans le cadre des travaux de biologie moléculaire, mais aussi et surtout pour sa bienveillance au quotidien. De l'autre côté de la frontière, je remercie également Pascal Dhulster, le directeur de l'Institut Charles Viollette, pour avoir accepté de m'accueillir au sein de l'institut et avoir supervisé l'avancement de ma formation doctorale. Merci également à François Coutte et Wolfgang Wohlleben pour avoir suivi l'avancement du projet et prodiguer de précieux conseils lors des différents comités de suivi de thèse.

Je remercie également les différentes personnes avec lesquels j'ai pu collaborer dans le cadre de ce projet et qui ont permis de faire avancer significativement mes travaux de recherche. Tout d'abord, un grand merci aux chercheurs des universités de Dresden, Tübingen et Wageningen, nos partenaires au sein du projet AstinProd, pour la qualité des discussions scientifiques et les soirées conviviales à chaque rassemblement biannuel. J'adresse un remerciement particulier à Thomas Schafhauser qui, durant une semaine, m'a consacré de son temps afin de me transmettre ses connaissances concernant la culture de notre endophyte bien aimé. Un grand merci également à Cédric Tarayre avec qui j'ai pu collaborer pendant plus de deux ans, et toujours dans la bonne humeur, sur la thématique de production homologue. Dans un autre registre, je remercie chaleureusement Loïc Meunier et Guillaume Brysbaert, mes

deux collaborateurs bio-informaticiens. Merci à Loïc pour son implication dans la rédaction de la partie « Genome mining » de la revue et nos nombreuses discussions enrichissantes autour de la bioinfo et des NRPSs. Merci à Guillaume pour sa disponibilité et tout ce qu'il a pu m'enseigner et m'apporter concernant la modélisation des protéines et l'approche de docking. Enfin, je remercie vivement les experts massistes qui m'ont assisté au cours du projet, notamment Anthony Arguelles Arias, Barbara Deracinois, et surtout mon frère de thèse Mickael Chevalier qui a souvent chercher désespérément de l'astine dans mes levures au cours des dernières années.

Toute ma reconnaissance s'adresse également aux personnes qui œuvrent chaque jour, que ce soit sur le plan administratif ou scientifique, pour nous permettre de réaliser nos travaux de recherches dans les meilleures conditions. Je pense notamment à Marina, Margue, Nuts, Martine, Danielle, Benoit, Jean-Rock et Samuel dans les locaux du Terra ; ainsi qu'à Cathy, Corine, Delphine, et Matthieu à l'ICV.

Je souhaite adresser un très grand merci aux personnes que j'ai côtoyé du côté de Gembloux. Le passage de Lille à Gembloux fut quelque peu dépaysant... Mais Gembloux est définitivement une ville à part, et on peut dire que les gembloutois ont l'art de faciliter l'acclimatation (même des français**!), tellement que j'y aurais volontiers posé mes valises plus longtemps. J'adresse un très sincère merci à l'ensemble des personnes travaillant au sein du bâtiment TERRA, qui par leur bonne humeur ont tous contribué à créer un cadre de travail convivial où j'ai pris plaisir à venir travailler ces dernières années. Je place une petite mention spéciale pour les membres de la team « Levure » supervisée par Patrick, et je remercie en particulier Olivia, Marie, Imen, et Diem pour les moments partagés au laboratoire ainsi que pour l'entraide au quotidien. Un grand merci également à mes compagnons de biologie moléculaire Nico, Seb, Greg, et Thibault. Au-delà du cadre du laboratoire, merci Greg pour ton soutien au quotidien, pour toute nos discussions centrées sur le foot (mais pas que !), et pour ton amitié. De même, j'adresse un immense merci à Seb... pour absolument tout. Tu as été un formidable soutien pour moi et un ami de confiance tout au long de la thèse, et je t'en suis sincèrement reconnaissant. J'en profite également pour adresser un big-up à Andrew, mon graphiste personnel, pour son aide sur la réalisation de la couverture du manuscrit et de la figure phare de la revue. La liste des remerciements commence maintenant à s'allonger dangereusement... Mais je souhaite adresser encore quelques remerciements à mes collègues Doc / Post-doc / Ingé de Gembloux : Sofija, Jelena, Ariane, Hannah, Clément, Romain, Adrien, Clotilde, Katrien, Hosni, et ceux que j'oublie... Merci à tous pour les échanges au labo, mais aussi pour les afterworks, les AGs, les Cénobs, les 6 heures, les Wallos, et tout ce qui a rythmé notre quotidien à Gembloux.

Passons maintenant au côté lillois... Je tiens tout d'abord à remercier l'ensemble du personnel de l'ICV, les ingés, techs, maître de conf, profs, etc., qui font vivre le laboratoire et nous permettent de réaliser nos travaux de recherche dans les meilleures conditions. Je suis en particulier très reconnaissant envers Qassim, Yazeen et Debarun pour tous les conseils avisés qu'ils ont pu me donner en biologie moléculaire sur les premiers mois où je suis arrivé à l'ICV. Je remercie également toutes les personnes

avec qui j'ai partagé les bureaux et qui ont su égayer mes journées, que ce soit dans le « grand » bureau (Kalim, Sandy, Alex Bricout, Alex Berth, Justine, Luiz, Alice, Quentin, ...), ou ces derniers mois dans le « petit » bureau (Adrien, Xavier, Dahlia, Morgane, Ramiro). Merci également à Marjo, Louis, Yannick, Paulo et l'ensemble de l'équipe Déino pour la bonne humeur qu'ils amènent dans le laboratoire et leurs compétences à la pétanque ! Enfin, je voudrais remercier en particulier ceux qui au fil des années sont devenus davantage des amis plutôt que des collègues du labo. Merci à Rémi, mon voisin de bureau pendant plus d'un an, avec qui j'ai partagé quelques matchs de foot, des soirées, et des chasses aux Pokémons. Merci à Maxime (et à Angelina, parce qu'ils vont souvent par paire) pour toutes ces soirées et ces moments de rigolades au labo et en dehors. Merci à Juliette qui a su plusieurs fois me remotiver, et qui par sa joie de vivre a illuminé quelques journées dans le bureau, et a endiablé quelques soirées à Solférino. Merci également pour son support sur la correction du manuscrit dans des délais record. Et pour finir, merci à Cyril, pour qui il me faudrait quelques pages des remerciements pour être exhaustif. Je vais me contenter de dire merci pour tout... Pour ton soutien scientifique et moral, pour les mois de colocs, pour les soirées sushi-geek, pour le trip à Rome, et pour ton aide précieuse sur la correction du manuscrit.

On l'oublie parfois, mais tout comme Rome, on ne se construit pas en un jour ! Je tiens donc à remercier quelques-uns de mes anciens mentors, que ce soit du monde académique ou industriel, qui m'ont porté jusqu'ici : Myriam Lefèvre, Alain Guiraud, Olivier van wuytswinkel, Éric Gontier, Claude Jamrozik, Esteban Freydell.

Je n'oublie pas non plus mes amis de longue date qui m'ont toujours soutenu, Nico, Mel, Edouard, Wiwi. Une mention spéciale à Mattou, mon voisin gembloutois chez qui j'ai passé bien des soirées. Je fais une dédicace à mes ex-binômes de TP, les premiers avec qui j'ai manipé, Nina et Anthony. Un immense merci à mon frerot Simon, pour son soutien indéfectible au jour le jour depuis des années (Zakwélie !).

Je vais terminer cette section en adressant mes plus profonds remerciements à ma famille. Tout d'abord, je ne pourrais jamais remercier assez mes parents pour le soutien qu'ils m'ont tous deux apportés tout au long de ma vie et de mes études. J'ajoute un grand merci à maman pour le temps qu'elle a consacré à la correction du manuscrit et à toute l'organisation autour de la soutenance. Je remercie du fond du cœur mes grands-parents qui par leur soutien présent ou passé m'ont donné la force de me surpasser et de tenir ma promesse de toujours continuer à travailler. Un immense merci à mes deux piliers, mes deux sœurs sur lesquelles je peux me reposer, Zélia et Lucile. Une petite dédicace à la team « cousin », Jojo, Edouard, Filou, Julien, et surtout Max qui m'a suivi dans mes délires footballistiques jusqu'au fin fond de la Russie. Je vais clore cette section de remerciements en remerciant du fond du cœur Barbara. Ton dévouement et ton soutien sans failles m'ont permis de mener ce projet à son terme malgré les aléas rencontrés en cours de route. Je te dois beaucoup, merci d'être toujours là pour moi au quotidien.

Pour conclure... Merci à ceux qui ont eu le courage de lire l'ensemble de cette section, et un plus grand merci encore à ceux qui oseront lire la suite.

Table des matières

Résumé	vii
Abstract	ix
Remerciements.....	xi
Table des matières	xv
Liste des figures	xxi
Liste des tableaux	xxix
Liste des abréviations	xxxix
Avant-propos	1
CHAPITRE 1 : Introduction bibliographique.....	5
1. Introduction	7
2. La voie de synthèse non-ribosomique	9
2.1. Définition et généralités	9
2.2. Organisation structurale et mode de fonctionnement des NRPSs	13
2.2.1. Description des domaines enzymatiques.....	13
2.2.2. Les synthétases : chaînes d'assemblage de peptides non-ribosomiques	21
2.3. Diversité de structures et d'activités des NRPs	23
3. Les astines : pentapeptides cycliques non-ribosomiques.....	25
3.1. Généralités et structures	25
3.2. Activités.....	27
3.3. Modes de production	30
3.3.1. Extraction à partir d' <i>Aster tataricus</i>	30
3.3.2. Production via l'endophyte <i>Cyanodermella asteris</i>	30
4. Stratégies d'optimisation de la production homologue de NRPs par un champignon filamenteux	32
4.1. Optimisation génétique de la souche	32
4.2. Optimisation des conditions de culture	34

5. Production hétérologue de peptides non-ribosomiques au sein d'usines cellulaires fongiques : de l'exploitation du génome à la production optimisée (Article)	35
5.1. Introduction	40
5.2. Computational tools and methods for genome mining	44
5.2.1. Recent advances in genome sequencing and annotation	44
5.2.2. Bioinformatics tools dedicated to NRPS study.....	45
5.2.3. Linking NRPS to related clusters or known products.....	49
5.2.4. Fungal NRPS analysis challenges	50
5.3. Heterologous production of NRPS in fungal host	53
5.3.1. Heterologous host choice.....	53
5.3.2. Genetic toolbox for the cloning of genes involved in NRP biosynthesis	57
5.3.3. Fungal expression system	61
5.3.4. Key factors to scale up heterologous production of NRPs	66
5.4. Penicillin heterologous production: a NRP model system	77
5.5. Concluding remarks.....	81
6. References	83

CHAPITRE 2 : Optimisation de la production homologue d'astine C par le champignon endophyte *Cyanoderrella asteris* (Article) 101

1. Introduction	107
2. Experimental section	109
2.1. Fungal strain	109
2.2. General cultivation conditions	109
2.3. Statistical methods.....	109
2.4. Immobilized-cell cultures	109
2.5. Effect of the inoculum type on immobilized-cell cultures of <i>C. asteris</i>	110
2.6. Support recycling procedure.....	110
2.7. Cultures in bioreactors.....	110
2.8. Astin C quantification.....	110
2.9. Scanning electron microscopy (SEM).....	111

3.	Results	112
3.1.	Optimization of the culture medium.....	112
3.2.	Comparison between two cultivation methods of <i>C. asteris</i> : planktonic and immobilized-cell conditions	113
3.2.1.	Macroscopic observations and characterization of the fungus through electron microscopy	113
3.2.2.	Biomass and astin C production by <i>C. asteris</i> in planktonic and ICC conditions.....	116
3.3.	Optimization of astin C production by <i>C. asteris</i> in immobilized-cell cultures	119
3.3.1.	Effect of the inoculum type on immobilized cell cultures of <i>C. asteris</i>	119
3.3.2.	Implementation of a production methodology: support recycling issues	121
3.3.3.	Optimization of immobilized-cell culture conditions to reach higher astin C yields	122
4.	Discussion	123
5.	References	126
CHAPITRE 3 : Décryptage de la voie de biosynthèse de l'astine par analyse bioinformatique		129
1.	Introduction	131
2.	Matériels et méthodes.....	133
2.1.	Analyse in silico des données génomiques.....	133
2.1.1.	Analyse globale du génome.....	133
2.1.2.	Analyse des différents gènes présents dans le cluster NRPS	134
2.2.	Prédiction de la structuration en modules / domaines des NRPSs	134
2.3.	Prédiction des peptides formés par analyse des domaines enzymatiques des NRPSs	134
2.3.1.	Prédiction de l'isomérisation des monomères.....	134
2.3.2.	Analyse de la spécificité des domaines d'adénylation	135
2.4.	Modélisation tridimensionnelle de la NRPS	135
2.4.1.	Établissement d'un modèle 3D de la NRPS	135

2.4.2.	Évaluation de la spécificité des domaines d'adénylation sur base du modèle tridimensionnel de la NRPS.....	136
3.	Résultats.....	137
3.1.	Analyse des données génomiques de <i>C. asteris</i>	137
3.1.1.	Recherche de clusters de gènes impliqués dans la synthèse de métabolites secondaires	137
3.1.2.	Organisation du cluster n° 6 (ctg45_orf8) de <i>C. asteris</i>	138
3.1.3.	Comparaison entre le cluster n° 6 de <i>C. asteris</i> et le cluster de gènes associé à la synthèse de la cyclochlorotine chez <i>T. islandicus</i>	140
3.2.	Analyse des séquences nucléotidique et protéique de AstS.....	141
3.2.1.	Prédiction de la structuration en modules / domaines	141
3.2.2.	Impact de la présence d'introns sur l'activité prédite des domaines	142
3.3.	Prédiction de la spécificité des domaines de la NRPS AstS.....	144
3.3.1.	Détermination du sous-type des domaines de condensation.....	144
3.3.2.	Prédiction de la spécificité de substrat des domaines d'adénylation	146
4.	Discussion.....	154
5.	Références	159
CHAPITRE 4 : Mise en place d'un système d'expression hétérologue chez la levure		163
1.	Introduction	165
2.	Matériels et méthodes	167
2.1.	Souches utilisées.....	167
2.1.1.	Souches de champignons filamenteux.....	167
2.1.2.	Souches de levures.....	167
2.1.3.	Souches de bactéries.....	167
2.2.	Conditions de cultures	167
2.2.1.	Culture des champignons filamenteux.....	167
2.2.2.	Culture des bactéries.....	168
2.2.3.	Culture des levures	168
2.3.	Extraction des ADN génomiques	168

2.3.1.	Méthode classiquement utilisée.....	168
2.3.2.	Mise au point d'une méthode d'extraction pour l'ADN de <i>C. asteris</i>	169
2.4.	Isolement du gène codant pour l'astine synthétase.....	169
2.4.1.	Par restriction de l'ADN génomique	169
2.4.2.	Par amplification PCR.....	170
2.5.	Analyse de l'adaptabilité des codons chez les levures	170
2.6.	Assemblage des plasmides	171
2.6.1.	Plasmides et amorces utilisés	171
2.6.2.	Méthodes communes à l'assemblage de l'ensemble des plasmides	174
2.6.3.	Différentes étapes nécessaires à l'obtention des plasmides.....	177
2.7.	Obtention des transformants de levure	180
2.7.1.	Transformation chez <i>S. cerevisiae</i>	180
2.7.2.	Transformation chez <i>Y. lipolytica</i>	181
2.8.	Séquençages et analyses	182
2.9.	Évaluation de la transcription (RT-qPCR)	182
2.10.	Analyses protéomiques.....	184
2.10.1.	Préparation des lysats protéiques.....	184
2.10.2.	Gel de protéines SDS-Page	185
2.10.3.	Purification des protéines sur colonne	186
2.10.4.	Western Blot.....	186
2.10.5.	Identification des protéines par spectrométrie de masse	187
2.11.	Synthèse in vitro du dipeptide 2Abu-Pro	187
3.	Résultats	190
3.1.	Usage des codons pour l'expression de <i>astS</i> chez les levures sélectionnées.....	190
3.2.	Extraction de l'ADN de <i>C. asteris</i> et isolement du gène codant pour la NRPS AstS	192
3.2.1.	Mise au point d'une méthode d'extraction de l'ADN génomique de <i>C. asteris</i>	192
3.2.2.	Obtention de la séquence <i>astS</i>	193

3.3.	Expression hétérologue de <i>astS</i> (17 kb) chez <i>S. cerevisiae</i>	196
3.3.1.	Obtention des différents transformants de <i>S. cerevisiae</i>	196
3.3.2.	Séquençage des clones positifs sélectionnés	202
3.3.3.	Évaluation de la transcription	205
3.3.4.	Analyses protéomiques	207
3.4.	Expression hétérologue de AstS tronquée (2 modules et C _{term}) chez <i>Y. lipolytica</i>	209
3.4.1.	Préparation des plasmides et obtention des transformants	209
3.4.2.	Évaluation de la transcription	212
3.4.3.	Analyses protéomiques	213
3.4.4.	Évaluation de la production du NRP : L-2Abu-LPro	216
4.	Discussions	218
5.	Références	222
CHAPITRE 5 : Discussion générale et Perspectives		227
1.	Discussion générale	229
2.	Perspectives	238
2.1.	Relatives à l'approche de production homologue	238
2.2.	Relatives au décryptage des voies de biosynthèse des NRPs par analyses bioinformatiques à partir de génomes fongiques	239
2.3.	Relatives à la production hétérologue de l'astine	239
3.	Références	241
Communications scientifiques		246

Liste des figures

Chapitre 1

Figure 1 : Représentation schématique d'une NRPS à 3 modules, assemblant un tri-peptide non-ribosomique. (A- domaine d'adénylation, T- domaine de thiolation, C- domaine de condensation, Te- domaine thioestérase, Mm- monomère)9

Figure 2 : Différentes étapes de la voie de biosynthèse de la tyrocidine A chez *B. brevis*. **A.** Présentation de l'opéron contenant les différents gènes codant pour les protéines composant la NRPS (*tycA*, *tycB*, *tycC*). **B.** Présentation des ARNms, obtenus après transcription des gènes NRPS, en cours de traduction. **C.** Représentation des 3 portions de NRPS correctement repliées et interconnectées par la présence de domaines COM. La tyrocidine A est assemblée selon un mode de synthèse linéaire grâce à l'action conjointe des domaines A (adénylation), T (thiolation), C (condensation), E (épimérisation), et Te (thioestérase).....11

Figure 3 : Différentes étapes de la voie de biosynthèse du ferrichrome chez *S. pombe*. **A.** Présentation du cluster de gènes contenant le gène codant pour la synthétase (*Sib1*), et celui codant pour une L-ornithine-N₅-oxygénase (*Sib2*) impliquée dans la synthèse du précurseur L-AHO* (L-N₅-acétyl-N₅-hydroxy-ornithine). **B.** Présentation de l'ARNm, obtenu après transcription de *Sib1*, en cours de traduction. **C.** Représentation de la NRPS comportant deux domaines d'adénylation fonctionnels (A₁ et A₃) et un inactif (A₂), ainsi que 6 domaines T (thiolation) et C (condensation). Le ferrichrome est le résultat d'un assemblage itératif où les 3 premiers domaines T sont chargés avec une glycine par le domaine A₁, et les 3 derniers avec L-AHO* par le domaine A₃. La libération et la cyclisation du NRP est, comme souvent chez les organismes fongiques, assurée par un domaine de condensation terminal (C_{term})...12

Figure 4 : Réaction d'adénylation activant le substrat préalablement spécifiquement reconnu au sein d'un domaine d'adénylation de NRPS.15

Figure 5 : Fonctionnement du domaine de thiolation. **A.** Schéma de la réaction de phosphopantéthéinylation permettant l'ajout du groupement 4'-phosphopantéthéine (PPant), obtenu à partir du coenzyme A, sur le domaine de thiolation. **B.** Réaction de thiolation avec fixation de l'amino-acyl préalablement adénylé sur la fonction thiol de la PPant.17

Figure 6 : Établissement de la liaison peptidique entre deux monomères, activés par deux modules successifs de la NRPS, au niveau du domaine de condensation (domaine A : adénylation, domaine T : thiolation, PPant : 4'-phosphopantéthéine transférase).....18

Figure 7 : Analyse des propriétés des peptides d'origine non-ribosomique répertoriés dans la base de données Norine en termes de structures (**A**) et d'activités (**B**). Certains NRPs présentant des activités multiples ont été pris en compte dans plusieurs des catégories représentées sur le diagramme B.....23

Figure 8 : Structure des différentes formes d'astine isolées à partir des racines d'*Aster tataricus*. **A.** Organisation générale des astines et éléments de différenciation entre les différentes formes : R₁, R₂, R₃, R₄, R₅, et liaison X---Y. **B.** Structure des formes d'astine J (linéaire), O, et P présentant certaines variations vis-à-vis du schéma décrit sur le panneau A (mis en évidence par la couleur brune).25

Figure 9 : Illustrations relatives à l'isolement de l'endophyte *Cyanodermella asteris* à partir de la plante *Aster tataricus*. Les différentes illustrations ont été adaptées à partir des figures provenant de l'étude de Jahn *et al.* **A.** Observation visuelle de l'inflorescence d'*A. tataricus*. **B.** Plantule d'*A. tataricus* stérile se développant sur un milieu contenant des hormones de croissance. **C.** Aspect de la plantule après 1 mois de culture, et premières étapes du développement du champignon endophyte. **D.** Développement précoce de *C. asteris* isolé sur un milieu agar pomme de terre (PDA). **E.** Aspect de *C. asteris* après 55 jours de culture sur milieu PDA.26

Figure 10: **A.** Profil de production des différentes formes d'astine produites par *Cyanodermella asteris* en culture pendant 28 jours en milieu PDB et MEB. **B.** Structure moléculaire de l'astine C, principale forme produite par *C. asteris*, comprenant 5 monomères : deux L-2-aminobutyrate (2Abu), une L-β-phénylalanine (βPhe), une L-cis-3,4-dichloroproline [Pro(Cl₂)] et une L-serine (Ser).27

Figure 11 : Différents effets observés *in vitro* chez des cellules hépatiques humaines L-02 après traitement avec des concentrations variables d'astine B. Figure adaptée à partir des résultats obtenus par Wang *et al.* (2014). **A.** Ratio de Bax / Bcl-2 mesuré par densitométrie après 24 h de traitement. **B.** Quantité de cytochrome C présent au sein des mitochondries et dans le cytosol, mesurée via un kit ELISA après 24 h de traitement. **C.** Activités enzymatiques des caspases-3 et -9 mesurées après 24 h de traitement. **D.** Mesure de la présence de composés induisant un stress oxydatif ou au contraire à caractère antioxydant. Le taux de dérivés réactifs de l'oxygène (en gris foncé) a été mesuré via un composé spécifique fluorescent (DCFH-DA) après 12 h de traitement. Le taux intracellulaire du composé antioxydant, le glutathion (en gris clair), a été déterminé après 24 h de traitement.....28

Figure 12 : Effets observés chez les cellules ganglionnaires (lymphocyte T) de souris soumises à un traitement de concentrations variables d'astine C pendant 24 h. Figure adaptée à partir des résultats obtenus par Shen *et al.* (2011). **A.** Mesure via la méthode MTT de la prolifération des cellules T actives ou inactives après traitement à l'astine C. **B.** Activités enzymatiques des caspases-3 et -9 mesurées après traitement à l'astine C.29

Figure 13: Multiple steps involved in NRP biosynthesis pathway. **A.** Transcription of the NRPS encoding gene. **B.** Translation of the NRPS mRNA. **C.** Folding of the NRPS in multi-enzymatic domains modular organization and addition of phosphopantetheine arm (*holo*-NRPS). Several domains are presented: A: adenylation; T: thiolation; C: condensation; C_{term}/T_e: thioesterase or terminal condensation domain (only found in fungal NRPSs); and an example of a potential additional epimerization (E) domain (converting an L-form amino acid to its D-form). **D.** Operating way of NRPS assembly line to generate NRP (Mm: monomer). *Epimerization reaction can occur either before or after condensation reaction.....41

Figure 14: Simplified workflow for novel NRP discovery by genome mining with an overview of the main bioinformatics tools available.....46

Figure 15: Analysis of the 50 most frequent monomers constituting the nonribosomal peptides classified in the Norine database. Each bar is subdivided into colored parts corresponding to the taxonomic clade from which the resulting NRPs was identified (i.e., white: bacteria, dark grey: fungi). The monomers specifically activated by A-domains which display pHMMs signatures enabling specificity predictions are highlighted in green.....51

Figure 16: Description of the four main cloning methods enabling the reconstitution, from multiple amplified PCR fragments, of long-length NRPS sequences into a targeted vector: **A.** Overlap extension PCR approach, **B.** Gibson assembly, **C.** USER Cloning, and **D.** TAR Cloning (homologous recombination).58

Figure 17: Overview of the limiting steps of NRP heterologous production and of the main solutions to overcome them and consider upgrading production rates.67

Figure 18: **A.** Presentation of the gene cluster involved in the biosynthesis of penicillin in *Penicillium chrysogenum*. **B.** Schematic representation of the penicillin G biosynthetic pathway in *P. chrysogenum*. ACV is assembled by the ACVS (encoded by the *pcbAB* gene), before to be modified into isopenicillin N by IPNS (encoded by *pcbC* gene), subsequently converted into penicillin G through combined activities of IAT and PCL (respectively encoded by *penDE* and *pclA* gene). *PclA** gene is not part of the penicillin biosynthetic cluster.70

Chapitre 2

Figure 1: Effect of medium modifications on *C. asteris* biomass and astin C production. Cultivations were performed at 22 °C in planktonic flasks under shaking at 150 rpm. Mean values and standard deviations were calculated from the data obtained from triplicate flasks filled with 50 mL of one of the following medium: **A.** Culture medium (malt extract 10 g/L, meat peptone 1 g/L, trace elements) containing either 10 or 20 g/L of one of the sugars: sucrose (Suc), mannose (Man), or glucose (Glc). Relative astin C quantity was expressed in percentage of the value obtained with 20 g/L of glucose (control condition). **B.** Culture medium (malt extract 10 g/L, glucose 10 g/L, meat peptone 1 g/L, trace elements) supplemented with the astin C building blocks: L-proline (Pro) 0.5 g/L, L-serine (Ser) 0.5 g/L, L-2-aminobutyric acid (2Abu) 1 g/L, L-β-phenylalanine (β-Phe) 0.5 g/L, or a combination of all four amino acids. Relative astin C quantity was expressed in percentage of the value obtained in the unsupplemented medium (control condition).112

Figure 2: Scanning electron microscopy (SEM)-views of *C. asteris* grown during 14 days in the MEAb medium (25 °C-160 rpm), as biofilm on the steel support (**A-I**) and as pellets in planktonic cultures (**J-O**). The following elements are indicated by a dedicated arrow: local bundle arrangements in microfibrils (dashed arrow), filamentous extracellular polymeric substances (solid arrow), spore-forming structures (dotted arrow), and hyphae showing branching (lozenge shaped arrow head). **A.** General views showing grid bars that have been sectioned and are covered by massive strands of filaments. **B, C.** Views of the external part of the biofilm. **D.** View of the middle part of the biofilm. **E, F, G.** Views of the deepest part of

the biofilm attached to the steel grid support. **H, I.** Details of the mycelium filaments (hyphae) showing branching and terminal ends. **J.** General view of a star-shaped pellet (a central core and several arms) formed in PLs. **K.** View of the middle part of an arm formed by a bundle of roughly parallel filaments. **L.** View of the middle part of arms. **M.** Details of the arm filaments with branching hyphae, local irregular cellular patterns and frequent characteristic fructifications in the middle part of arms. Spores exhibit a characteristic round-shaped terminal end and a constricted junction with the mother-hyphae. **N.** View of the central core pellet with a regular terminal end, terminal spores and abundant EPS. **O.** Detailed view of the hyphae showing branching and EPS interlacing the hyphae.114

Figure 3: Comparison between two cultivation methods of *C. asteris*: ICCs (light grey bars) and PL (dark grey bars) in the MEAb or the d5MEAb. Cultivations were performed at 25 °C in flasks under shaking at 160 rpm. Mean values and standard deviations were calculated from the data obtained from triplicate flasks filled with 100 mL of culture medium. **A, B.** Follow-up of the reducing sugars consumption, quantified via the DNS method, in ICCs and PLs in (A) d5MEAb medium or (B) MEAb medium. **C, D.** Follow-up of the pH evolution of (C) d5MEAb medium or (D) MEAb medium while growing *C. asteris* in ICCs and PLs. **E, F.** Follow-up of the astin C concentration (mg/mL) in the ICCs and PLs of *C. asteris* in (E) d5MEAb medium or (F) MEAb medium. **G, H.** Mean value of astin C yield ($Y_{P/S}$) in mg per gram of consumed reducing sugar in time, with ICC and PL conditions, in (G) d5MEAb medium or (H) MEAb medium.117

Figure 4: Evaluation, in ICCs, of the effect of some cultivation parameters on the production of astin C and the development of *C. asteris*. **A, B.** Influence of the inoculum type on astin C production and *C. asteris* growth. Cultivations were performed under shaking at 160 rpm, at 25 °C, in flasks filled with 100 mL of the MEAb medium. Mean values and standard deviations were calculated from the data obtained from triplicate flasks inoculated either with 1 mL of hyphae from a 15 day-old culture (mycelium) or 1 mL of a mixture of hyphae and spores obtained from a 55 day-old culture (spores). **A.** The fungal final dry biomass (g/L) attached to the support (dark grey) or released into the medium (light grey). **B.** The follow-up of the astin C concentration (mg/mL). **C, D.** Evaluation of the impact of the washing treatment applied to the support on *C. asteris* attachment and development (C) and on astin C production (D). The supports used to immobilize the cells were previously washed either with a treatment A (H₂O only), a treatment B (H₂O followed by 24 h immersion in NaOH 20%), or a treatment C (identical to treatment B with an additional step of 1 h immersion in 100 °C H₂O). Cultivations were performed under shaking at 160 rpm, at 25 °C, in flasks filled with 100 mL of the MEAb medium inoculated with spores. Mean values and standard deviations were calculated from the data obtained from triplicate flasks. **E.** Influence of the medium composition on the astin C production in ICCs. Cultivations were performed under shaking at 160 rpm, at 25 °C, in flasks filled with 100 mL of culture medium inoculated with spores. Mean values and standard deviations were calculated from the data obtained from triplicate flasks filled either with the MEAb or MEAc culture medium. **F.** Evaluation of the development of *C. asteris* on the stainless steel support in two different pH cultivation conditions. *C. asteris* was grown as a biofilm during 36 days at 25 °C in immobilized-cell condition, in DASGIP bioreactors with a pH controlled at 4.8 or without any pH control. The visual aspects of the stainless steel supports obtained after cultivation of *C. asteris* with the pH controlled at 4.8, and without the pH control, are compared with the visual aspect of the support without cells.120

Chapitre 3

Figure 1 : Organisation modulaire des NRPSs, responsables de l'assemblage de la Cct et HCct, encodées par *cctN* et *pisl3812-04907* chez *T. islandicus*.131

Figure 2 : Différentes étapes d'analyses bioinformatiques (tenant compte des spécificités relatives aux séquences fongiques) pour l'identification *in silico*, à partir des données génomiques de *C. asteris*, de la NRPS d'astine.133

Figure 3 : Organisation du cluster n° 6 (*ctg45_orf8*) de *C. asteris* contenant le gène *astS* correspondant potentiellement à la synthétase d'astine. Les rôles supposés, dans la biosynthèse de l'astine, des protéines encodées par les différents gènes du cluster ont été évalués via les outils BLASTp et BLASTn.138

Figure 4 : Analyse comparative des gènes présents au sein du cluster n° 6 (contenant le gène *astS* codant potentiellement pour la NRPS de l'astine) chez *C. asteris* et ceux identifiés dans le cluster de gènes responsable de la synthèse de la cyclochlorotine *T. islandicus*. Le pourcentage d'homologie entre les protéines encodées par certains gènes des deux clusters sont indiquées.140

Figure 5 : Analyse de la séquence protéique de la NRPS encodée par *AstS* via PKS/NRPS Analysis et InterProScan.141

Figure 6 : Séquences protéiques des domaines d'adénylation des modules 1 et 3 obtenues après épissage ou non des séquences identifiées comme introns. Les signatures caractéristiques sont mises en évidence par la couleur rouge. Les signatures caractéristiques localisées dans les zones annotées comme introns sont présentées en vert.143

Figure 7 : Portion de l'arbre phylogénétique généré via NaPDoS, mettant en évidence la conservation des séquences protéiques entre les domaines de condensation composant AstS, les domaines C_{term} T.isl1 et T.isl2, et les domaines répertoriés dans la base de données. ...145

Figure 8 : Structures des deux acides aminés : L-cystéine et L-Sérine.....147

Figure 9 : Superposition des conformations tridimensionnelles de : **A.** 1AMU (violet) et 5T3D (cyan), RMSD = 1.462 Å ; **B.** 1AMU (violet) et 3ITE (jaune), RMSD = 1.275 Å...149

Figure 10 : Résultat de la modélisation du module 2 de l'AstS via le logiciel phyre2 (Visualisation PyMOL) avec en rouge le domaine d'adénylation, en bleu celui de thiolation, et en vert celui de condensation.150

Figure 11 : Visualisation via PyMOL du positionnement des différents ligands au sein des domaines d'adénylation durant les simulations de « docking ». **A.** Validation de l'approche de docking sur la structure cristallisée de 1AMU en présence du cofacteur AMP. La position théorique du résidu Phe (en rose) (obtenu par « docking ») est comparée à la position expérimentale attribuée au substrat Phe (en rouge). **B.** Positionnement de l'AMP (vert), de la leucine (bleu) et de la proline (rouge) au sein du domaine d'adénylation du module 2 de l'AstS modélisée. **C.** Positionnement de l'AMP (vert), de l'Asp (bleu) et de la β-Phe (rouge) au sein du domaine d'adénylation du module 5 de l'AstS modélisée.151

Figure 12 : Proposition de schéma détaillant la voie de biosynthèse des astines C, F et G. La synthétase de 5 modules est codée par le gène *astS*, puis activée par l'action d'une phosphopantéthéine transférase dont le gène n'a pas été identifié. Les gènes *astU* et *astV* sont potentiellement impliqués dans la synthèse des monomères précurseurs de l'astine (respectivement β -Phe et L-2Abu). Une fois assemblé, le peptide est libéré et cyclisé par l'action du domaine C_{term}, générant ainsi le NRP d'astine sous la forme G. Les formes d'astine F et C (mono- et di-chlorées) sont éventuellement obtenues par l'action d'une halogénase qui n'a pas été identifiée au sein du cluster de gènes n° 6. Le gène *astT* (transporteur de type ABC) est ensuite potentiellement impliqué dans la sécrétion du NRP.....157

Chapitre 4

Figure 1 : Séquence chimérique générée pour le clonage dans *Y. lipolytica*, composée d'un fragment de la séquence de *astS* codant pour les deux premiers modules, directement suivis, dans le même cadre de lecture ouvert (ORF), par la séquence codant pour le domaine de condensation terminal. La synthétase tronquée produite, si elle est fonctionnelle, doit assembler et libérer un dipeptide L-2Abu – L-Pro.166

Figure 2: Protocole appliqué pour générer un surnageant et un produit de synthèse *in vitro* du dipeptide L-2Abu – L-Pro à partir d'une culture de RIY261 et de B3R2.189

Figure 3 : Nombre de codons du gène *astS* présentant une adaptabilité relative inférieure à 20 % chez *Y. lipolytica*, *P. pastoris*, *H. polymorpha*, et *S. cerevisiae*, déterminé par l'intermédiaire de Graphical Codon Usage Analyser.190

Figure 4 : Impact de différents pré-traitements de lyse cellulaire sur la quantité d'ADN génomique de *C. asteris* obtenu après utilisation des kits Qiagen, Norgen Plant, ou Norgen soil. Chaque extraction a été réalisée en duplicata et les quantités d'ADN sont exprimées en ng par mg de biomasse fongique utilisée lors de l'extraction.192

Figure 5 : Position des sites de restriction reconnus par les enzymes *FseI* (rose), *SgrDI* (vert) et *EcoRI* (blanc) au sein de la région génomique (54 kb) où se situe le gène *astS*....193

Figure 6 : Profil de migration obtenu avant et après restriction de l'ADNg de *C. asteris* en utilisant les enzymes de restriction *SgrDI* (A) et *EcoRI* (B).194

Figure 7 : Mise au point d'un protocole pour l'amplification et la purification du gène *astS* (17 kb). **A.** Résultat de l'amplification en utilisant différentes polymérases avec le couple d'amorces *AstS1/AstS2*. **B.** Résultat de l'amplification en utilisant la GoTaq long PCR Mastermix avec les couples d'amorces *AstS1/AstS2* et *AstS2/AstR3-tag-stop*, après mise au point d'un protocole de migration sur gel d'agarose adapté aux fragments de grande taille.....195

Figure 8 : Différentes étapes nécessaires à l'assemblage du plasmide pESC-ura-sfp-Ast5'-Ast3', avec le clonage en premier lieu du gène *sfp* amplifié par PCR à partir de l'ADN de *B. subtilis* (A), puis de l'insertion du fragment *Ast5'* amplifié par PCR à partir de l'ADN de *C. asteris* (B), suivie par le clonage du fragment *Ast3'* également amplifié par PCR à partir de l'ADN de *C. asteris* (C).197

Figure 9 : Différentes étapes nécessaires à l'assemblage du plasmide pYX112-Ast5'-Ast3', avec l'insertion en premier lieu du fragment *Ast5'* amplifié par PCR à partir de l'ADN de *C. asteris* (A), suivi par le clonage du fragment *Ast3'* également amplifié par PCR à partir de l'ADN de *C. asteris* (B).198

Figure 10 : Plasmide pESC-sfp-AstS.RESTR généré par TAR cloning dans *S. cerevisiae* par co-transformation du plasmide pESC-sfp-5'-3' et de l'ADNg de *C. asteris* restreint par *EcoRI*. Les positions des sites de restriction reconnus par les enzymes *FspAI* et *PacI* utilisées pour libérer par la suite le fragment afin de l'insérer dans le plasmide pYX112Ast5'-Ast3' sont indiquées.....200

Figure 11 : Plasmide pYX112-AstS.RESTR généré par TAR cloning dans *S. cerevisiae* par co-transformation du plasmide pYX112-Ast5'-Ast3' et du fragment de 17 kb obtenu par restriction *FspAI* / *PacI* du plasmide pESC-sfp-AstS.RESTR.201

Figure 12 : Stratégie de clonage développée chez *S. cerevisiae*, avec l'insertion de *astS* (obtenu par restriction ou par amplification PCR de l'ADNg de *C. asteris*) au sein du plasmide pESC-ura (promoteur inductible GAL1 / GAL10), puis le transfert de cette séquence de pESC-ura à pYX112 (promoteur constitutif TPI – Triose Phosphate Isomérase).....202

Figure 13 : Visualisation de la position de la mutation N°9 (Ser3170Pro) au sein du domaine de condensation du module 4 et estimation de l'impact de cette mutation sur la structure du domaine via le logiciel Misense3D. A. Vue à l'échelle des liaisons atomiques. B. Vue de la structure tridimensionnelle et de l'organisation en hélices α / feuillets β204

Figure 14 : Valeurs moyennes relatives de $2^{-\Delta\Delta Ct}$ obtenues pour les échantillons provenant de la souche DS.Py.62 en fixant les échantillons provenant de la souche DS.PE.Ec73 en tant que condition contrôle, à 16 h de culture (A) et à 36 h de culture (B). Les moyennes et écarts types ont été calculés sur des cultures réalisées en duplicata.207

Figure 15 : Profil de migration des protéines totales extraites à partir des cellules de DSy5, DS.Pe.Ec73, et DS.Py62, après migration dans un gel SDS-Page. Les marqueurs de poids moléculaire sont constitués de thyroglobuline (669 kDa), de ferritine (440 kDa), de catalase (232 kDa), et de lactate déshydrogénase (140 kDa).208

Figure 16 : Assemblage des séquences codantes pour les 2 premiers modules de l'AstS, suivis par le domaine C_{term}, au sein du plasmide RIP191-URA3. A. Gel d'électrophorèse après migration du fragment de 6012 pb amplifié par PCR avec le couple d'amorces *BamHI-Ast6kb-fw* / *Ast6kb-sfiIA-SfiIB-AvrII*. B. Carte plasmidique de RIP191-URA3-Ast6kb après insertion du fragment de 6 kb amplifié par PCR. Les sites de restrictions *SfiI* requis pour l'insertion du domaine C_{term} sont localisés au niveau de l'extrémité 3' du fragment Ast6kb. C. Gel d'électrophorèse après migration du fragment de 1229 pb amplifié par PCR avec le couple d'amorces *SfiA-Cterm-fw* / *SfiB-Cterm-rv*. D. Carte plasmidique de RIP191-URA3-Ast6kb-Cterm après insertion du fragment de 1,2 kb amplifié par PCR.210

Figure 17 : Obtention du plasmide RIP222-LEU2-NpgA par ligation entre RIP222-LEU2 et le fragment PCR NpgA de 1059 pb (restreint par l'action des enzymes *Hind*III et *Avr*II dont l'emplacement des sites de restriction est indiqué).212

Figure 18 : Profil de migration sur SDS-Page des lysats protéiques extraits à partir des cultures de RIY261 (contrôle négatif) et B3R2 (souche transformée *astS.Tr+*, *npgA+*) après 8 et 24 h de culture sur milieu YPD supplémenté en erythritol.....214

Figure 19 : Révélation du western blot réalisé avec des anticorps dirigés contre le tag 6x HIS sur des protéines extraites à partir des souches B3R2 et RIY261 cultivées sur YPD ou YNB pendant 8 ou 24 h.....215

Figure 20 : Mise au point d'une méthode de dosage pour le dipeptide L-2Abu – L-Pro (A) par l'intermédiaire d'une chromatographie en phase liquide (spectre UV à 215 nm) (B) couplée à un spectromètre de masse (C).216

Chapitre 5

Figure 1 : Description des deux approches (homologue et hétérologue) développées au cours de cette étude afin d'accroître les rendements de production en astine C.230

Liste des tableaux

Chapitre 1

Tableau 1 : Séquences signatures caractéristiques de certains des domaines enzymatiques composant les NRPS (adénylation, thiolation, condensation, thioestérase, méthylation, et épimérisation)	14
Table 2: List of condensation domain subtypes and their associated reactions.....	48
Table 3: Previous studies reporting NRP heterologous production in yeasts	54
Table 4: Previous studies reporting NRP heterologous production in filamentous fungi...56	

Chapitre 2

Table 1: Mean values of the final weighted dry biomass, the biomass yield $Y_{X/S}$, and the specific astin C production $Y_{P/X}$ in planktonic and immobilized-cell conditions.....	117
---	-----

Chapitre 3

Tableau 1 : Nombre de clusters détectés lors de l'analyse antiSMASH du génome de <i>C. asteris</i> , répartis selon le type de métabolites secondaires associés à ces clusters.	137
Tableau 2 : Annotation détaillée des domaines composant les NRPSs identifiées dans les 5 clusters détectés par antiSMASH dans le génome de <i>C. asteris</i>	137
Tableau 3 : Résultat de l'analyse réalisée avec NapDoS pour les 5 domaines de condensation identifiés au sein de la NRPS AstS, et les 2 domaines C_{term} des NRPSs de Cct et HCct (T.isl1 et T.isl2).	144
Tableau 4: Prédiction de la spécificité des domaines d'adénylation de la NRPS AstS via NRPS Predictor2 (np : non prédit).	146
Tableau 5 : Domaines / modules de PKS / NRPS cristallisés constituant des templates possibles pour la modélisation des différents modules de la NRPS AstS. (Structures d'origine eucaryote surlignées en bleu, et procaryote en blanc).	148
Tableau 6: Scores obtenus pour le « docking » des 20 acides aminés protéinogènes au sein de 1AMU (domaine A cristallisé).....	152

Chapitre 4

Tableau 1 : Plasmides utilisés ou assemblés au cours de cette étude.	172
Tableau 2 : Séquences des différentes amorces utilisées au cours de cette étude. Les sites de restriction et les tag 6xHIS ajoutés par le biais des amorces sont respectivement indiqués en orange et en vert.	173
Tableau 3 : Conditions générales des amplifications PCR réalisées avec l'ADN polymérase Q5® High-Fidelity DNA Polymerase en termes de composition du mélange (A) et de réaction réalisée (B).	174
Tableau 4 : Conditions générales des amplifications PCR réalisées avec la Taq DNA polymérase en termes de composition du mélange (A) et de réaction réalisée (B).	175
Tableau 5 : Liste et séquences des amorces utilisées dans les réactions qPCR pour l'évaluation de la transcription des différents gènes.	183
Tableau 6 : Nature et nombre d'occurrence des codons présentant une adaptabilité relative pour l'expression du gène <i>astS</i> inférieure à 20 % chez <i>S. cerevisiae</i> et <i>Y. lipolytica</i>	191
Tableau 7 : Résultats obtenus après TAR cloning chez <i>S. cerevisiae</i> , en co-transformant avec une longue séquence <i>astS</i> isolée par des approches différentes.	199
Tableau 8 : Résultats du séquençage du gène <i>AstS</i> inséré dans le plasmide pESC-sfp- <i>AstS</i> .PCR. 13 mutations ont été détectées au sein de la séquence génétique dont deux mutations silencieuses (vert) et deux susceptibles d'impacter fortement la structure tridimensionnelle par inclusion de l'acide aminé cyclique proline (rouge).	203
Tableau 9 : Résultats obtenus en termes de Ct (cycle seuil) lors des qPCR réalisées, afin d'évaluer la présence du transcrit du gène <i>astS</i> , avec le couple d'amorces <i>qpAstS1</i> / <i>qpAstS2</i> sur les ADNc synthétisés à partir des ARNm extraits des souches DSy5 (contrôle négatif), DS.Pe.Ec73, et DS.Py.62.	206
Tableau 10 : Résultats obtenus en termes de Ct (cycle seuil) lors des RT-qPCR réalisées à partir des ADNc extraits des souches B3R2 et RIY261 en utilisant le couple d'amorces <i>qpAend1</i> / <i>qpAend2</i> (A) et <i>qp-npg1-fw</i> / <i>qp-npg1-rv</i> (B).	213
Tableau 11 : Fragments peptidiques appartenant à la séquence protéique de <i>AstS</i> .Tr identifiés par spectrométrie de masse. (M* : oxydation +15,99 sur le résidu M).....	215

Liste des abréviations

ABC	Transporteur type « ATP Binding Cassette »
ACV	d-(L- α -aminoadipyl)-L-cysteinyl-D-valine
ADN	Acide désoxyribonucléique
ADNc	Acide désoxyribonucléique complémentaire
ADNg	Acide désoxyribonucléique génomique
ARNm	Acide ribonucléique messenger
BCG	Cluster de gènes biosynthétiques
BSA	Albumine de sérum bovin
Cct	Cyclochlorotine
D-	Dextrogyre
EPS	Substances polymériques extracellulaires
Fw	« Forward »
HCct	Hydroxy-cyclochlorotine
ICC	Culture de cellules immobilisées
IC50	Concentration inhibitrice médiane
L-	Lévogyre
L-AHO	L-N ₅ -acétyl-N ₅ -hydroxy-ornithine
MFS	Transporteur type « Major Facilitator Superfamily »
Mm	Monomère
MW	Masse moléculaire
NCBI	« National Center for Biotechnology Information »
NRP	Peptide non-ribosomique (« NonRibosomal Peptide »)
NRPS	Synthétase de peptide non-ribosomique (« NonRibosomal Peptide Synthetase »)
ORF	Cadre de lecture ouvert
PDB	Protein Data Bank
pHMMs	« Profile hidden markov models »
PKS	Synthase de polykétide
PL	Culture planctonique
PM	Marqueur de poids moléculaire
PPant	4'-phosphopantéthéine
PPTase	Phosphopantéthéine transférase
qsp	Quantité suffisante pour
Rv	« Reverse »
SGS	Séquençage seconde génération
SMs	Métabolites secondaires
TGS	Séquençage troisième génération
TPI	Triose phosphate isomérase

• Acides aminés et monomères

Abu	Acide aminobutyrique
Aib	Acide α -amino-iso-butyrique

Ant	Acide anthranilique
Asp	Acide aspartique
Cys	Cystéine
Dhb	Acide 2,3-dihydroxybenzoïque
Glu	Acide glutamique
Gln	Glutamine
Gly	Glycine
His	Histidine
Leu	Leucine
Lys	Lysine
Orn	Ornithine
Phe	Phénylalanine
Pro	Proline
Pro(Cl ₂)	L-cis-3,4-dichloroproline
Ser	Sérine
Trp	Tryptophane
Val	Valine

• **Principaux composés chimiques**

AcN	Acétonitrile
AMP	Aminoacyl-adénosine monophosphate
ATP	Aminoacyl-adénosine triphosphate
CHCl ₃	Chloroforme
Cl	Chlore
DCFH-DA	2',7'-Dichlorofluorescin di-acétate
dATP	Désoxy adenine tri-phosphate
dNTP	Les désoxy-nucléotides tri-phosphate
DTT	Dithiothréitol
EDTA	Ethylène diamine tétra-acétate
IAA	Acide 3-indole acétique
IPTG	Isopropyl-β-D-thiogalactopyranoside
MeOH	Méthanol
MgCl ₂	Dichlorure de magnésium
Mg ²⁺	Ion magnésium
NaCl	Chlorure de sodium
NH ₂	Groupement amine
PPi	Pyrophosphate
SO ₂	Dioxyde de soufre
SO ₂ -NH	Groupement sulfonamide
SDS	Dodécylsulfate de sodium
TFA	Acide trifluoroacétique
Tris HCl	Trisaminométhane hydrochloride
X-Gal	5-bromo-4-chloro-3-indolyl β-galactopyranoside

• Méthodes

ELISA	Dosage d'immunoabsorption par enzyme liée (« Enzyme-linked immunosorbent assay »)
LC-MS	Chromatographie liquide suivie par spectrométrie de masse
MS	Spectrométrie de masse
MWCO	Seuil de coupure (« Molecular weight cut-off »)
PCR	Réaction en chaine par polymérase
qPCR	Réaction en chaine par polymérase quantitative
RMN	Résonance magnétique nucléaire
TAR Cloning	« Transformation associated recombination cloning »
USER Cloning	« Uracil specific excision reagent cloning »
SEM	« Scanning electron microscopy »
SDS Page	Électrophorèse sur gel polyacrylamide et dodécylsulfate de sodium
UF	Ultrafiltration

• Synthétases de peptides non-ribosomiques

ACVS	Synthétase de d-(L- α -aminoadipyl)-L-cysteinyl-D-valine
AstS	Synthétase d'astine
DhbE	(2,3-dihydroxybenzoyl) adenylate synthase
EntF	Entérobactine synthase (module F)
LgrA	Synthétase de gramicidine linéaire
TqaA	Synthétase de fumiquinazoline F
TycA	Synthétase de tyrocidine (module A)
TycC	Synthétase de tyrocidine (module C)
VibH	Synthétase de vibriobactine
VinN	Enzyme d'adénylation dans la biosynthèse de vicenistatine

• Domaines des synthétases de peptides non-ribosomiques

Domaine A	Domaine d'adénylation
Domaine C	Domaine de condensation
Domaine ^L C _L	Domaine de condensation liant un acide aminé forme L à un acide aminé forme L
Domaine ^D C _L	Domaine de condensation liant un acide aminé forme D à un acide aminé forme L
Domaine C/E	Domaine de condensation avec une activité d'épimérisation
Domaine C _{start}	Domaine de condensation « starter »
Domaine C _{term}	Domaine de condensation terminal
Domaine COM	« Acceptor communication-mediating domain »
Domaine E	Domaine d'épimérisation
Domaine M	Domaine de méthylation
Domaine PCP	Domaine « Peptidyl carrier protein »
Domaine Re	Domaine réductase
Domaine T	Domaine de thiolation
Domaines Te	Domaine thioestérase

• **Paramètres et unités**

°C	Degrés Celsius
Å	Ångström
Ct	Cycle seuil
Da	Dalton
DO	Unité de densité optique
g	Gramme
h	Heure
kb	Kilobase
kDa	Kilodalton
L	Litre
M	Mole/L
mg	Milligramme
min	Minute
mL	Millilitre
mM	Millimole/L
mV	Millivolt
ng	Nanogramme
nm	Nanomètre
nM	Nanomole/L
pb	Paire de base
pH	Unité de potentiel hydrogène
rpm	Rotation par minute
RMSD	Root Mean Square Deviation
sec	Seconde
T _{élong}	Temps d'élongation
T _m	Température d'hybridation
µg	Microgramme
µL	Microlitre
µm	Micromètre
µM	Micromole/L
V	Volt

Avant-propos

Les propriétés des microorganismes ont été exploitées pendant des siècles afin d'élaborer des produits alimentaires tels que le fromage, le pain, le vin, et la bière. Plus récemment, les applications possibles se sont diversifiées avec par exemple l'utilisation de produits naturels d'origine microbienne en tant que composés pharmaceutiques, biocarburants, bio-pesticides, ou bio-surfactants. Le marché global des produits microbiens représente aujourd'hui 153,6 milliards d'euros, et devrait atteindre 221,1 milliards d'euros d'ici 2023 avec une croissance annuelle estimée de 8,7 % (Source : Microbial Products Market Research Report, 2019). Les composés présentant une activité relative au domaine pharmaceutique (antibiotique, anti-inflammatoire, anticancéreux, antiallergique, antiviral, immunosuppresseur) constituent la plus grande part de ce marché (39 %). On estime que la moitié des composés à applications médicales approuvés entre 1981 et 2010 sont des produits d'origine naturelle. Par exemple, pour le marché des antibiotiques, les pénicillines et céphalosporines, composés naturels de type β -lactamines, représentent à eux seuls 47 % de ce marché.

La mise à disposition sur le marché pharmaceutique d'un nouveau produit est le résultat d'un processus long et coûteux, puisqu'il faut compter en moyenne 10 à 15 ans et plus d'un milliard d'euros investis de l'étape de recherche du composé d'intérêt jusqu'à sa commercialisation. Seuls 0.01 % des produits naturels actifs caractérisés aboutiront finalement à une mise sur le marché. Afin de répondre aux attentes d'un marché en expansion perpétuelle tout en considérant l'investissement nécessaire en matière de temps et d'argent, deux principaux défis sont à relever. Le premier comprend le développement de méthodes facilitant le processus de criblage nécessaire à la découverte de nouveaux produits microbiens bioactifs. Le second, qui en découle, consiste, lorsqu'un composé prometteur est identifié, à intensifier les efforts afin d'optimiser sa production d'une part, et d'évaluer son activité et sa toxicité d'autre part.

C'est dans ce contexte et avec l'ambition de relever ces défis que s'inscrit ce travail de thèse. Pour cela, les travaux se sont concentrés sur la molécule d'astine, dont les propriétés bénéfiques ne sont plus à démontrer puisque celle-ci est utilisée indirectement depuis des décennies en médecine traditionnelle chinoise et japonaise. L'astine C présente notamment des activités anti-tumorale et anti-inflammatoire qui sont prisées par l'industrie pharmaceutique. Récemment, dans le cadre d'un projet européen (ASTINPROD project - ERA-IB-15-039), dans lequel s'inscrivent ces travaux, le microorganisme fongique responsable de la biosynthèse de l'astine a été identifié et isolé. Cette découverte a ouvert de nouvelles perspectives concernant la production d'astine à plus grande échelle. L'objectif de ce projet de thèse consiste à étudier la voie de biosynthèse de l'astine chez ce champignon afin de développer et d'optimiser la production de ce composé de manière homologue ou hétérologue.

Pour répondre aux besoins du projet, la collaboration et l'expertise de deux structures ont été nécessaires : l'équipe MiPi (« Microbial Processes and Interactions »), incluse dans le centre de recherche et d'enseignement TERRA en

Belgique, et l'Institut Charles Viollette (ICV) en France. L'équipe MiPi possède des compétences avancées en matière de production de composés d'origine microbienne et de caractérisation de leurs activités ; tandis que l'ICV est spécialisée dans la production de nouveaux peptides et lipopeptides d'origine non-ribosomique via l'utilisation et le développement d'outils bioinformatiques, l'optimisation des souches à l'échelle moléculaire, et l'élaboration de procédés de production. Cette collaboration s'est concrétisée par la mise en place d'un laboratoire international associé MOM (« Metabolites d'Origine Microbienne »), dédié au criblage à haut débit de nouveaux métabolites microbiens secondaires, de la mise en évidence de leurs activités biologiques à l'optimisation de leur production. Les travaux menés lors de ce projet de thèse ont été cofinancés dans le cadre de ce LIA par les universités de Lille et de Liège-Gembloux Agro-Bio Tech, et ont bénéficié par ailleurs du support et de l'expertise de l'unité de Microbiologie et Génomique de Gembloux.

Ce projet de thèse s'est articulé autour de 4 objectifs majeurs qui vont chacun donner lieu à un chapitre du manuscrit de thèse. Tout d'abord une étude bibliographique a été réalisée afin de compiler l'étendue actuelle des connaissances sur les astines et leur mode de synthèse de type non-ribosomique. Cette étude a permis la rédaction d'une revue détaillant les différentes étapes à considérer pour implémenter une production hétérologue dans un hôte fongique, de l'identification des gènes à l'optimisation de la production en passant par le choix de la souche. Le second objectif consiste à accroître la production homologue d'astine C, notamment en optimisant les conditions de culture de son producteur naturel. Le troisième objectif, développé en parallèle, visait à décrypter la voie de biosynthèse de l'astine chez le champignon producteur. Dans ce but, une analyse bioinformatique du génome séquencé de ce champignon a été réalisée. Une fois la voie de biosynthèse identifiée, le dernier objectif consistait à la transférer dans un hôte hétérologue de type levure afin de mettre en place un système de production d'astine avec un microorganisme plus facilement manipulable et cultivable.

De ce fait, l'organisation du manuscrit de thèse se présentera comme suit :

- **Chapitre 1 : Introduction bibliographique**

Cette section a pour objectif d'une part de définir le mode de synthèse non-ribosomique, de compiler les données relatives à l'étude des astines et d'identifier les différentes stratégies envisageables pour optimiser une production homologue ou développer une production hétérologue de peptides d'origine non-ribosomique chez les organismes fongiques. Cette dernière partie a fait l'objet d'une revue de littérature comportant une présentation des outils bioinformatiques relatifs à l'identification des gènes, des problématiques relatives au choix de l'hôte, ainsi que des différentes étapes à considérer afin de parvenir à exprimer ces gènes chez l'hôte, puis à optimiser la production hétérologue. Cette analyse sera présentée en anglais telle qu'elle a été publiée dans le journal *Biotechnology Advances*.

- **Chapitre 2 : Optimisation de la production d'astine C par le champignon endophyte *Cyanoderrella asteris* en conditions planctoniques et cellules immobilisées (Article)**

Ce chapitre est consacré à l'optimisation des conditions de culture du champignon *C. asteris*, producteur naturel de l'astine C. En particulier, deux systèmes de fermentation ont été étudiés et comparés. Le premier permet le développement du champignon sous forme de pelotes en condition planctonique, tandis que le second, en condition de cellules immobilisées, favorise le développement d'un biofilm. Les travaux relatifs à ce chapitre ont fait l'objet d'une publication scientifique. Les résultats obtenus seront présentés en anglais tels qu'ils ont été publiés dans le journal *Biotechnology Journal*.

- **Chapitre 3 : Décryptage de la voie de biosynthèse de l'astine par analyse bioinformatique**

Ce chapitre présente les différents outils et analyses bioinformatiques ayant permis l'identification des gènes impliqués dans la biosynthèse des astines chez le champignon producteur.

- **Chapitre 4 : Mise en place d'un système d'expression hétérologue chez la levure**

Cette section détaille les différentes étapes mises en œuvre pour transférer la voie de biosynthèse de l'astine C de *Cyanoderrella asteris* chez les levures *Saccaromyces cerevisiae* et *Yarrowia lipolytica*. Cette partie comprend la mise au point des méthodes d'extraction et d'amplification de l'ADN de *C. asteris*, les travaux de biologie moléculaire ayant permis d'aboutir à des transformants de levures, ainsi que les analyses réalisées sur les mutants obtenus.

- **Chapitre 5 : Discussion générale et perspectives**

Les résultats présentés dans les quatre chapitres précédents seront discutés dans leur ensemble au sein de cette section afin d'apporter une conclusion et de mettre en lumière les perspectives à venir concernant ce travail de recherche.

Chapitre 1

Introduction bibliographique

1. Introduction

Les métabolites secondaires, bien que n'étant pas décrits comme tels, ont été utilisés indirectement depuis des siècles afin de formuler des produits alimentaires, des colorants, des parfums, des produits cosmétiques, des médicaments, ou même des poisons. Dans un premier temps, ce sont les composés naturels d'origine végétale qui ont été principalement étudiés et utilisés [1]. La morphine, un alcaloïde issu de *Papaver somniferum*, est l'un des premiers composés bioactifs à avoir été isolé et caractérisé [2]. Les premiers travaux relatifs aux métabolites secondaires d'origine microbienne sont généralement attribués au professeur H. Raistrick et à son équipe qui ont caractérisé plus de 200 métabolites issus d'organismes fongiques à partir de 1923 [3]. L'intérêt pour les produits naturels microbiens a ensuite été décuplé à la suite de la découverte de la très célèbre pénicilline, aux propriétés antibiotiques, produite par *Penicillium notatum* [4]. D'autres composés antibiotiques de première importance ont ensuite été isolés entre 1953 et 1970 : vancomycine, norcardicine, imipénem, et aztréonam [5]. Les années qui ont suivi ont été marquées par la caractérisation de nombreux métabolites secondaires présentant des propriétés pharmaceutiques intéressantes : les statines régulatrices du cholestérol [6], la cyclosporine permettant de prévenir le risque de rejet lors d'une greffe d'organe [7], la griséofulvine aux propriétés antifongiques [8], la cytochalasine E à l'effet antitumoral [9], l'acide mycophénolique présentant une activité immunosuppressive [10], la myriocine dont le dérivé est utilisé pour lutter contre la sclérose en plaque [11]. A l'heure actuelle, on estime que plus de 50 000 composés naturels ont été isolés à partir de microorganismes.

La distinction entre métabolites primaires et secondaires est basée sur plusieurs critères. En outre, les métabolites secondaires sont généralement produits uniquement durant une certaine période du cycle cellulaire (souvent à l'issue de la phase de croissance), et ne sont pas indispensables à la croissance et la survie de la cellule. Ce sont des composés présentant très souvent une activité biologique et dont la production est habituellement spécifique à un organisme ou à un groupe taxonomique très restreint [12]. La très grande diversité de structures des métabolites secondaires rend difficile la classification de ces composés. La première option consistait à classer ces métabolites en fonction de la principale source de carbone utilisée lors de leur synthèse : métabolites obtenus à partir (1) d'acides aminés, (2) de composés acétates, ou (3) de sucres [1]. Plus récemment, l'élucidation et la compréhension d'un très grand nombre de voies de biosynthèse des métabolites secondaires ont permis de développer une classification basée sur les classes d'enzymes impliquées dans leur synthèse [13]. Selon cette classification, les principales classes de composés sont constituées par : (1) les terpènes synthétisés par les terpènes cyclases, (2) les polykétides assemblés à partir d'acétyl- et malonyl- coenzyme A par des polykétide synthases, (3) les peptides non-ribosomiques résultant de l'assemblage entre autres d'acides aminés protéinogènes et non-protéinogènes par des complexes multienzymatiques appelés synthétases peptidiques non-ribosomiques.

La voie de synthèse non-ribosomique constitue le principal sujet d'étude de cette synthèse bibliographique. Cette voie a été pour la première fois mentionnée en 1965, avant d'être décrite six ans plus tard, avec l'identification de la synthétase de la gramicidine produite par *Bacillus subtilis* [14,15]. Auparavant, un autre composé dérivé de la voie non-ribosomique avait été isolé sans que la voie de biosynthèse ne soit identifiée : la pénicilline [4]. Le précurseur de la pénicilline, un tripeptide, est devenu au fil des années l'un des peptides non-ribosomiques les plus étudiés. Aujourd'hui, d'autres composés d'origine non-ribosomique sont considérés notamment pour leurs propriétés thérapeutiques. C'est le cas de l'astine, un pentapeptide cyclique présentant une activité anticancéreuse, initialement détecté chez *Aster tataricus* et dont la production a été récemment attribuée à *Cyanodermella asteris* [16]. Ce peptide non-ribosomique a été sélectionné comme principal modèle d'étude dans le cadre de ce projet de doctorat et fera donc l'objet d'un état de l'art exhaustif. Pour finir, en tenant compte des propriétés particulières de la voie de synthèse non-ribosomique, différentes solutions seront proposées afin d'une part, de faciliter l'identification de ces voies au sein des génomes des microorganismes, et d'autre part, de transférer ces voies de biosynthèse au sein d'un hôte hétérologue parfois plus adapté en matière de production.

2. La voie de synthèse non-ribosomique

2.1. Définition et généralités

Les peptides d'origine non-ribosomique, comme leur nom l'indique, ne sont pas obtenus par la voie classique de synthèse des protéines (voie ribosomique), mais assemblés par des synthétases peptidiques non-ribosomiques. Ces synthétases et peptides associés sont communément nommés respectivement NRPS et NRP en référence à leur dénomination en anglais « Non-Ribosomal Peptide Synthetase » et « Non-Ribosomal Peptide ». Les NRPS s'organisent en macro-complexes subdivisés en modules comportant de multiples domaines enzymatiques, dont les activités combinées permettent l'assemblage de peptides non-ribosomiques [17] (**Figure 1**).

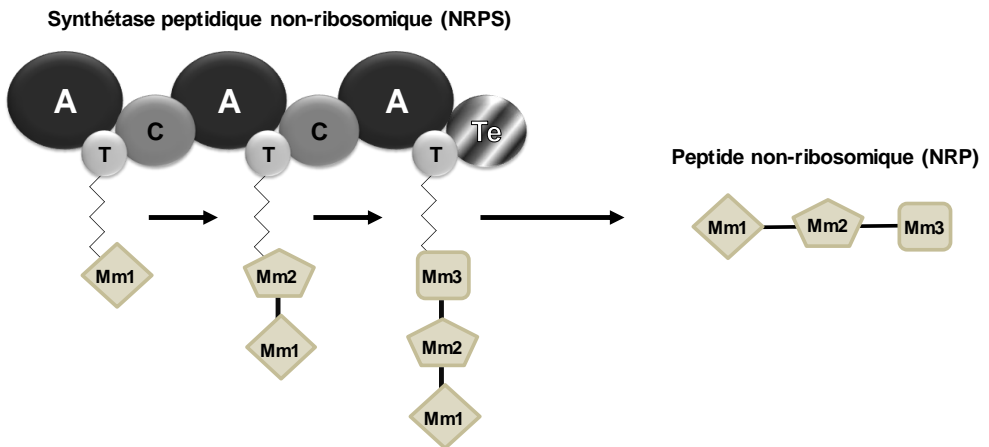


Figure 1 : Représentation schématique d'une NRPS à 3 modules, assemblant un tri-peptide non-ribosomique. (A- domaine d'adénylation, T- domaine de thiolation, C- domaine de condensation, Te- domaine thioestérase, Mm- monomère).

Il existe d'autres complexes enzymatiques assemblant des peptides tels que les synthétases de cyanophycine, les glutathionne synthases, les D-alanine-D-alanine ligases, et les L-amino acide α -ligases. Mais au contraire des NRPSs, l'activation du substrat ne se fait pas via un groupement aminoacyl-adénosine monophosphate (AMP), mais via un groupement aminoacyl phosphate [18].

La présence de gènes impliqués dans la production de peptides non-ribosomiques a été mise en évidence chez de nombreux organismes provenant de genres taxonomiques variés. La très grande majorité des NRPs caractérisés proviennent de bactéries ou de champignons filamenteux [19]. Néanmoins l'existence de cette voie de synthèse non-ribosomique a également été mise en évidence chez les archées [20], et les cyanobactéries [21]. La synthèse de ferrichrome par *Schizosaccharomyces pombe* est à ce jour le seul cas reporté de synthèse non-ribosomique chez une levure [22]. En revanche, aucune production avérée de peptides non-ribosomiques au sein d'organismes eucaryotes appartenant aux clades taxonomiques metazoa ou plantae n'a pu être mise en évidence.

Les gènes impliqués dans la biosynthèse d'un même NRP sont très souvent regroupés au sein d'un même cluster de gènes chez les organismes fongiques, ou opéron chez les organismes bactériens. La NRPS peut être soit codée par un unique gène comme par exemple la ferrichrome synthétase chez *S. pombe* [23] soit par plusieurs gènes comme pour la tyrocidine synthétase chez *Bacillus brevis* [24] (**Figure 2A** et **Figure 3A**). Dans ce second cas, le NRP assemblé résulte de l'activité d'une NRPS constituée de plusieurs protéines. La longueur des gènes codant pour les NRPSs peut atteindre plus de 10 kilobases selon la taille de la synthétase et le nombre de modules qu'elle comporte. En moyenne, chaque module correspond à une séquence codante de 3,5 kb. Outre les gènes codant pour la synthétase, d'autres gènes impliqués dans la voie de biosynthèse du NRP sont très souvent présents au sein du cluster / opéron. Les gènes présents correspondent généralement à des familles de transporteurs, à des enzymes impliquées dans la synthèse des monomères précurseurs du NRP, ou à des enzymes de décoration modifiant le NRP synthétisé. En cas de toxicité du peptide produit, des gènes conférant à l'organisme producteur une résistance vis-à-vis du NRP, peuvent également être présents au sein du cluster / opéron [25].

Les synthétases peptidiques non-ribosomiques sont produites suivant la voie classique de synthèse des protéines [26]. Les gènes NRPS sont dans un premier temps transcrits en ARNm par une ARN polymérase. La longue séquence d'ARNm est ensuite traduite en séquence protéique par les ribosomes (**Figure 2B** et **Figure 3B**), puis la macro-protéine suit le processus de maturation des protéines afin d'acquérir sa structure tertiaire. La NRPS acquiert alors sa structure multi-modulaire, ainsi que les multiples domaines enzymatiques requis pour l'assemblage de peptides non-ribosomiques (**Figure 2C** et **Figure 3C**). Certaines études ont montré une localisation des NRPSs à proximité de la membrane cytoplasmique des cellules. C'est notamment le cas de la NRPS de pyoverdine chez *Pseudomonas aeruginosa* [27], et de la synthétase d'entérobactine chez *Escherichia coli* [28]. Cette localisation périmembranaire des NRPSs peut s'expliquer par le fait que la plupart des NRPs produits sont ensuite sécrétés dans le milieu extracellulaire.

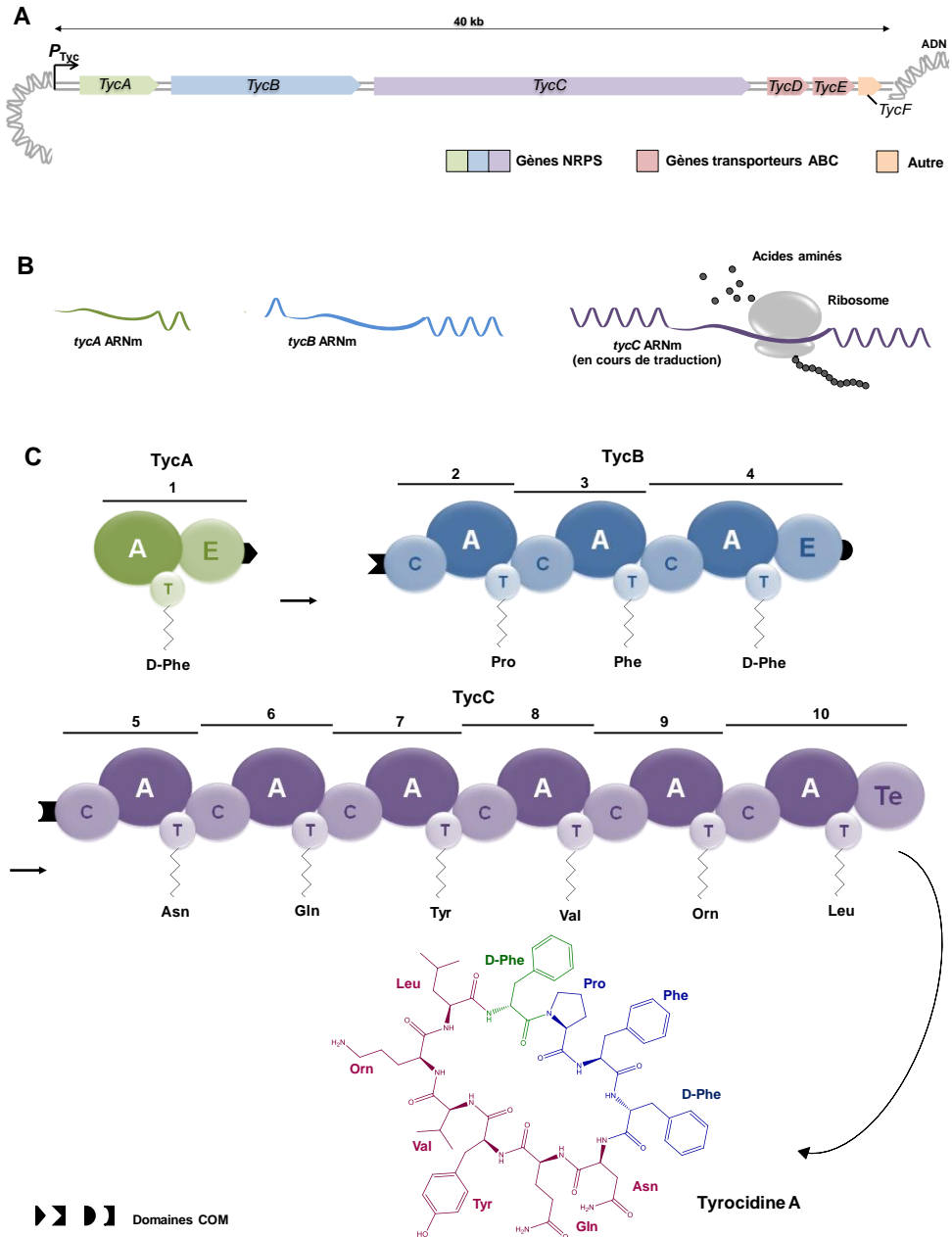


Figure 2: Différentes étapes de la voie de biosynthèse de la tyrocidine A chez *B. brevis*. **A.** Présentation de l'opéron contenant les différents gènes codant pour les protéines composant la NRPS (*tycA*, *tycB*, *tycC*). **B.** Présentation des ARNm, obtenus après transcription des gènes NRPS, en cours de traduction. **C.** Représentation des 3 portions de NRPS correctement repliés et interconnectées par la présence de domaines COM. La tyrocidine A est assemblée selon un mode de synthèse linéaire grâce à l'action conjointe des domaines A (adénylation), T (thiolation), C (condensation), E (épimérisation), et Te (thioestérase).

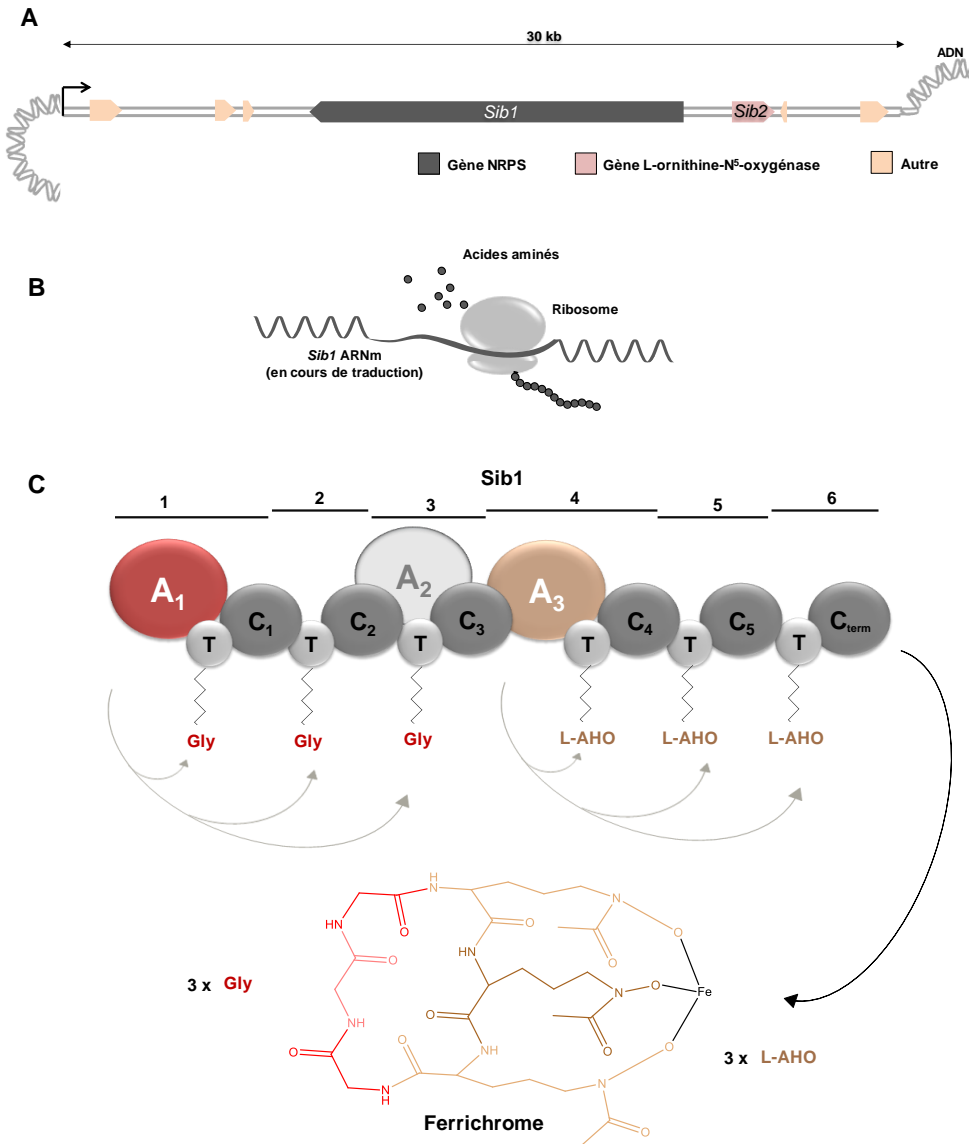


Figure 3 : Différentes étapes de la voie de biosynthèse du ferrichrome chez *S. pombe*. **A.** Présentation du cluster de gènes contenant le gène codant pour la synthétase (*Sib1*), et celui codant pour une L-ornithine-N₅-oxygénase (*Sib2*) impliquée dans la synthèse du précurseur L-AHO* (L-N₅-acétyl-N₅-hydroxy-ornithine). **B.** Présentation de l'ARNm, obtenu après transcription de *Sib1*, en cours de traduction. **C.** Représentation de la NRPS comportant deux domaines d'adénylation fonctionnels (A₁ et A₃) et un inactif (A₂), ainsi que 6 domaines T (thiolation) et C (condensation). Le ferrichrome est le résultat d'un assemblage itératif où les 3 premiers domaines T sont chargés avec une glycine par le domaine A₁, et les 3 derniers avec L-AHO* par le domaine A₃. La libération et la cyclisation du NRP est, comme souvent chez les organismes fongiques, assurée par un domaine de condensation terminal (C_{term}).

2.2. Organisation structurale et mode de fonctionnement des NRPSs

2.2.1. Description des domaines enzymatiques

Les synthétases de peptides non-ribosomiques sont des enzymes multi-modulaires, où chaque module est lui-même subdivisé en différents domaines enzymatiques. Ces domaines sont facilement identifiables grâce à l'existence de courtes séquences d'acides aminés, appelées signatures, qui sont très conservées d'une NRPS à une autre et d'un organisme à l'autre. Chaque domaine possède ses propres signatures caractéristiques (**Tableau 1**) [29]. Cette spécificité est notamment utilisée par certains outils bioinformatiques (ex. antiSMASH) pour prédire la structuration en domaines d'une NRPS donnée.

Tableau 1: Séquences signatures caractéristiques de certains des domaines enzymatiques composant les NRPS (adénylation, thiolation, condensation, thioestérase, méthylation, et épimérisation) [29].

Domaine	N° Signature	Séquence d'acides aminés consensus
Adénylation	A1	L(TS)Y _x EL
	A2	LKAG _x AYL(VL)P(LI)D
	A3	LAY _{xx} YTSG(ST)TG _x PKG
	A4	FD _x S
	A5	N _x YGPTE
	A6	GEL _x I _x G _x G(VL)ARGYL
	A7	Y(RK)TGDL
	A8	GR _x D _x QVKIRG _x RIELGEIE
	A9	LP _x YM(IV)P
	A10	NGK(VL)DR
Thiolation	T	D _x FF _{xx} LGG(HD)S(LI)
Condensation	C1	S _x AQ _x R(LM)(WY) _x L
	C2	RHE _x LRT _x F
	C3	MHH _x ISDG(WV)S
	C4	Y _x D(FY)AVW
	C5	(IV)G _x FVNT(QL)(CA) _x R
	C6	(HN)QD(YV)PFE
	C7	RD _x SRNPL
Thioestérase	Te	G(HY)S _x G
Méthylation	M1	VL(DE)G _x G _x G
	M2	NELS _x YRY _x AV
	M3	VE _x S _x ARQ _x G _x LD
Épimérisation	E1	PIQ _x WF
	E2	HH _x ISDG(WV)S
	E3	D _x LL _x A _x G
	E4	EGHGRE
	E5	RTVGWFT _{xx} YP(YV)PFE
	E6	P _{xx} G _x GYG
	E7	FNYLG(QR)

Parmi les différents domaines présents chez les NRPSs, 2 catégories peuvent être distinguées : les domaines principaux et les domaines secondaires. Les domaines d'adénylation, de thiolation, de condensation et de thioestérase, indispensables à la synthèse et la libération du NRP, constituent les domaines principaux. Les domaines secondaires sont impliqués dans la modification des monomères incorporés au cours de la synthèse non-ribosomique. Ils sont représentés notamment par les domaines de méthylation, d'épimérisation, de cyclisation, et d'oxydation [30].

2.2.1.1. Les domaines principaux

- **Domaine d'adénylation**

Les domaines d'adénylation sont des domaines très conservés chez l'ensemble des synthétases de NRPs. Leur taille varie de 500 à 600 acides aminés. Ils appartiennent à la superfamille des enzymes formant des adénylates [31]. Leur rôle dans la synthèse non-ribosomique est crucial puisqu'ils sont responsables de la sélection et de l'activation des monomères incorporés dans le peptide non-ribosomique. Bien que certains domaines d'adénylation soient capables de reconnaître plusieurs monomères différents [32], ces domaines présentent en règle générale, une spécificité très élevée pour un monomère en particulier [33].

Ces dernières années, la structure de plusieurs domaines d'adénylation a été cristallisée et résolue. C'est le cas notamment du domaine d'adénylation activant la phénylalanine au sein de la NRPS de la gramicidine S [34], et des domaines A présents au sein des modules NRPS de DhbE [35], de VinN [36], ou de TycA [37]. Ces structures cristallisées ont largement contribué à l'élucidation du mode de fonctionnement des domaines d'adénylation. Les domaines A s'organisent en deux sous-domaines, l'un plus court du côté C-terminal, et l'autre plus long où se situe le site actif, du côté N-terminal. Dans un premier temps, le monomère spécifiquement reconnu va se fixer dans la poche catalytique du domaine d'adénylation. Par la suite une molécule d'ATP est consommée entraînant la fixation de la fraction adénosine de l'AMP sur le monomère présent, générant ainsi un amino-acyl adénylate [38]. Le site actif comporte également un site de liaison compatible avec $MgCl_2$. Les ions Mg^{2+} constituent un cofacteur important, impliqué dans la neutralisation de la charge de l'ATP et du groupement PPi libre à l'issue de la réaction [39,40] (**Figure 4**).

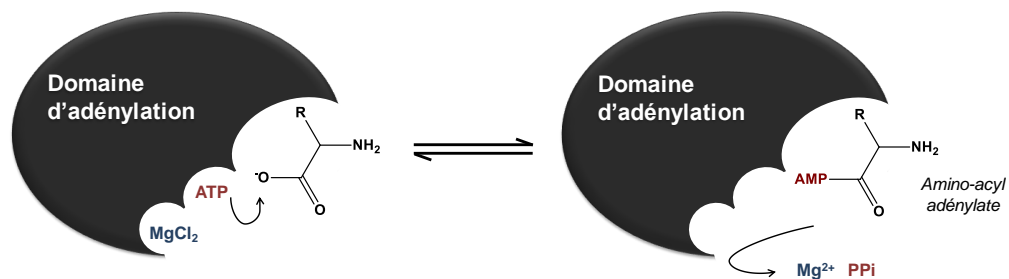


Figure 4 : Réaction d'adénylation activant le substrat préalablement spécifiquement reconnu au sein d'un domaine d'adénylation de NRPS.

Bien que les synthétases soient considérées comme des structures flexibles, les domaines d'adénylation présentent une certaine rigidité et une faible mobilité vis-à-vis des autres domaines tout au long de la synthèse non-ribosomique [35].

- **Domaine de thiolation**

Les domaines de thiolation, aussi appelés domaines PCP pour « peptidyl carrier proteins », sont de courts domaines composés d'environ 80-100 acides aminés [26]. Les substrats préalablement activés par les domaines d'adénylation sont fixés sur les domaines de thiolation. Cette fixation n'est possible qu'en présence d'un bras phosphopantéthéine terminé par un groupement sulfhydryle. Lorsque le bras phosphopantéthéine (PPant) est ajouté aux domaines de thiolation, la NRPS est convertie de la forme *apo*-NRPS à la forme active *holo*-NRPS [41]. La phosphopantéthéinylation des domaines de thiolation est une modification post-traductionnelle des NRPSs, réalisée par l'action d'une phosphopantéthéine transférase (PPTase). Cette enzyme catalyse la fixation de la partie 4'-phosphopantéthéine, provenant du coenzyme A, sur une sérine du domaine de thiolation. La sérine constituant le point de fixation est un résidu qui est conservé chez toutes les NRPSs et qui s'inclut dans la signature du domaine T (G-G-D/H-S-L) [42] (**Figure 5A**). Le degré de spécificité des PPTases varie d'une enzyme à l'autre, puisque certaines catalysent cette réaction de manière très spécifique sur certaines synthétases uniquement, tandis que d'autres comme la PPTase Sfp produite par *Bacillus subtilis* présentent une faible spécificité [43].

Certains domaines de thiolation, tels que ceux présents chez TycA et TycC (synthétase de la tyrocidine) ont été cristallisés [44,45]. Les substrats préalablement adénylés sont fixés aux domaines de thiolation au niveau du groupement sulfhydryle (site thiol) présent sur le bras phosphopantéthéine (**Figure 5B**). La liaison entre l'amino-acyl et l'atome de soufre entraîne la libération du groupement adénylate. Le complexe amino-acyl thioester ainsi formé, présente une très grande flexibilité vis-à-vis des autres domaines de la NRPS [30]. Aucune structure correspondant à une poche de liaison où pourrait se fixer le substrat n'a été mise en évidence dans les domaines de thiolation. Par conséquent, il est probable que ces domaines enzymatiques ne présentent aucune forme de spécificité pour les substrats fixés [46].

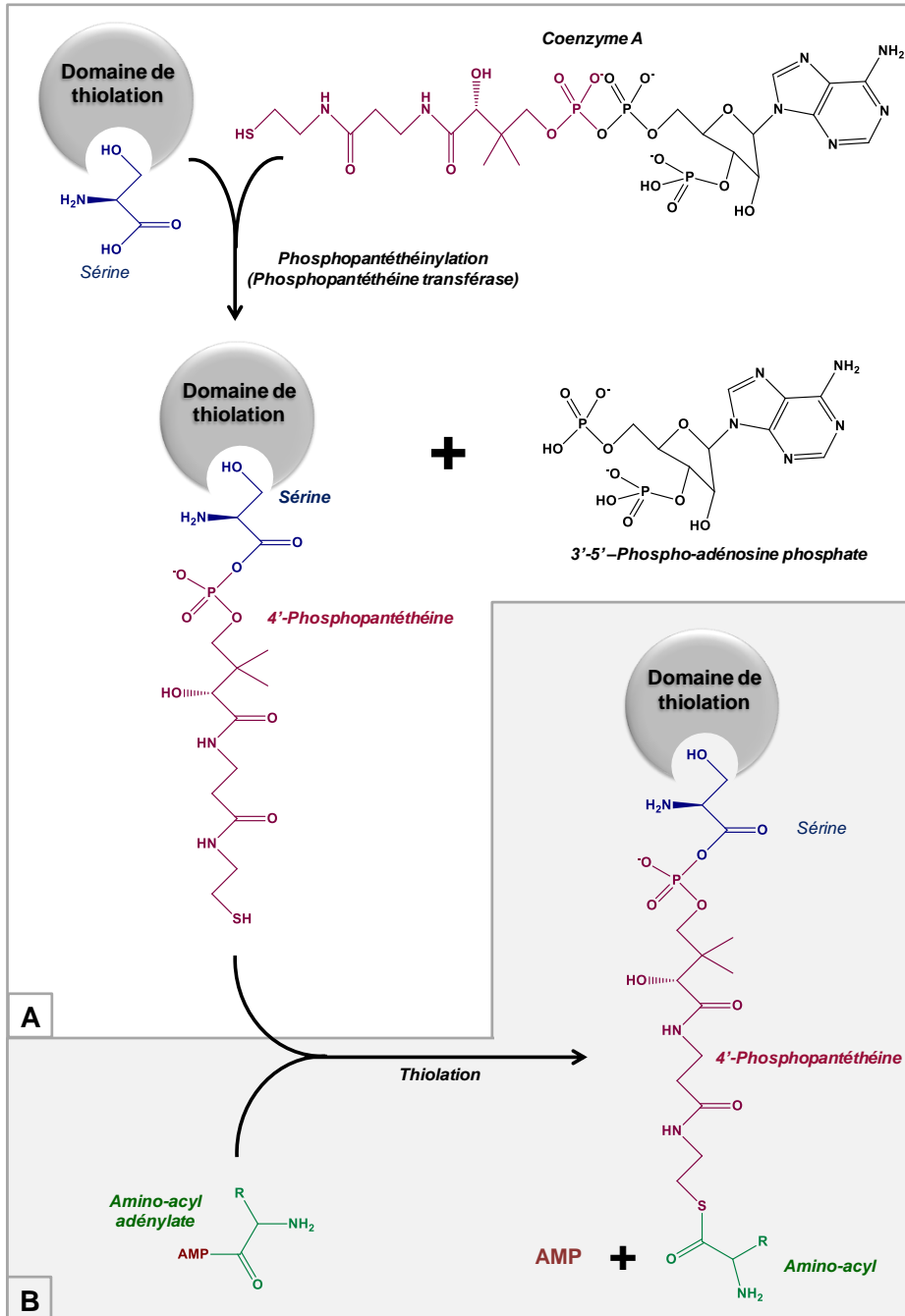


Figure 5 : Fonctionnement du domaine de thiolation. **A.** Schéma de la réaction de phosphopantéthéinylation permettant l'ajout du groupement 4'-phosphopantéthéine (PPant), obtenu à partir du coenzyme A, sur le domaine de thiolation. **B.** Réaction de thiolation avec fixation de l' amino-acyl préalablement adénylé sur la fonction thiol de la PPant.

- **Domaine de condensation**

Les domaines de condensation, composés en moyenne de 450 résidus, catalysent la formation de la liaison peptidique entre les différents monomères composant le peptide non-ribosomique [46]. Le domaine de condensation de la synthétase VibH provenant de *Vibrio cholerae* est le premier à avoir été cristallisé. Ce domaine est responsable de l'établissement de la liaison entre une norspermidine et un acide dihydroxybenzoïque durant la synthèse du sidérophore de vibriobactin [47]. Ce domaine s'organise en deux sous-domaines dont les structures sont très similaires. Ces deux sous-domaines correspondent à des sites de fixation permettant d'accueillir les monomères, portés par les domaines de thiolation de deux modules successifs, entre lesquels la liaison va s'établir (**Figure 6**). Une organisation similaire a également été observée lors de la cristallisation du domaine de condensation de la synthétase TqaA provenant de *Penicillium aethiopicum* [48].

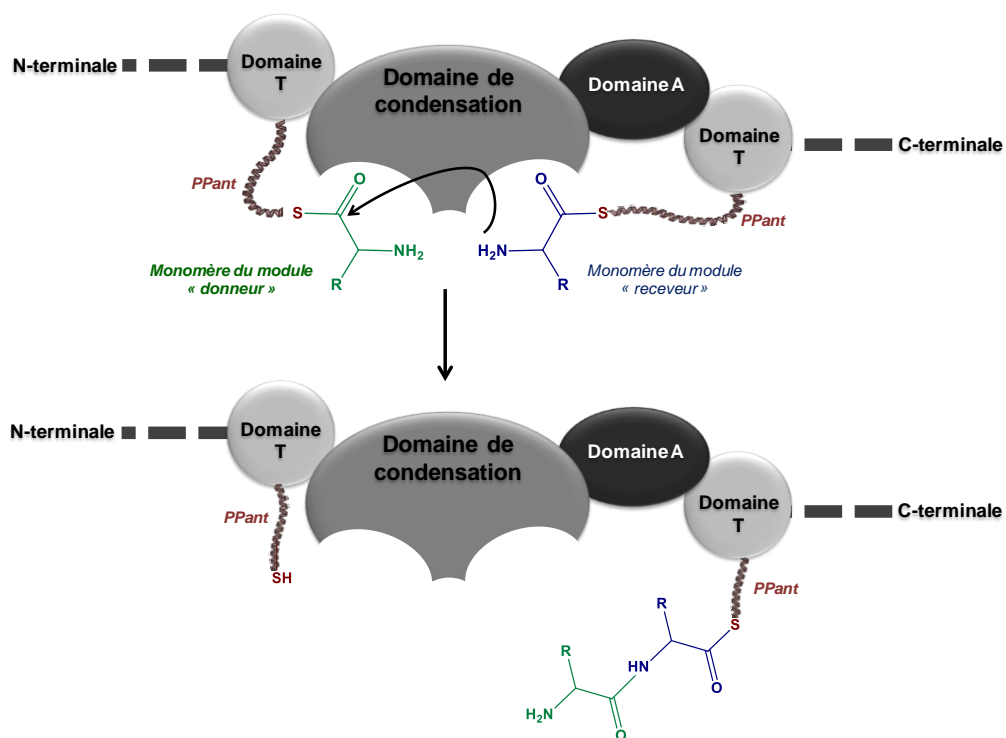


Figure 6 : Établissement de la liaison peptidique entre deux monomères, activés par deux modules successifs de la NRPS, au niveau du domaine de condensation (domaine A : adénylation, domaine T : thiolation, PPant : 4'-phosphopantéthéine transferase).

Ces sous-domaines possèdent, comme les domaines d'adénylation, une spécificité de substrat. Cette spécificité est plus marquée chez le sous-domaine situé du côté de la partie C-terminale, fixant le monomère « receveur » du monomère ou de la chaîne peptidique en cours de synthèse [49].

En règle générale, une fois les monomères fixés, on observe une dé-protonation du groupement NH_2 du monomère « receveur », situé dans le sous-domaine en C-terminal, suivi d'une attaque nucléophile sur le groupement carbonyle du monomère « donneur », situé dans le sous-domaine en N-terminal, générant la liaison peptidique (**Figure 6**) [31]. Le site catalytique des domaines de condensation est caractérisé par la présence de la séquence signature C3 (MHHxISDG[WV]S) [50]. En particulier, la seconde histidine (depuis la partie N-terminale) et l'acide aspartique ont été identifiés comme étant des résidus cruciaux dans le maintien de l'activité catalytique des domaines de condensation [51].

• **Domaine thioestérase / condensation terminal / réductase**

Les domaines thioestérase (Te) constituent bien souvent le domaine final des NRPSs, catalysant le clivage entre le peptide non-ribosomique néo-synthétisé et le domaine de thiolation du dernier module qui le porte. Ces domaines enzymatiques sont apparentés à la famille des α,β -hydrolases caractérisée notamment par une triade catalytique de trois résidus : une sérine, une histidine et un acide aspartique [38]. Le peptide non-ribosomique linéaire porté par le domaine de thiolation réagit avec la sérine du site actif du domaine Te, générant un intermédiaire peptidyl-O-Te. Une réaction d'hydrolyse, impliquant une attaque nucléophile de cet intermédiaire par un groupement hydroxyle, entraîne la libération du peptide non-ribosomique [52,53].

Dans certains cas, la libération du peptide non-ribosomique par l'action du domaine thioestérase est accompagnée d'une cyclisation de ce peptide. Ce mode de fonctionnement a été observé par exemple au sein des domaines Te des NRPSs de la surfactine [54] ou de la tyrocidine (**Figure 2C**) [55]. Dans le cas de la surfactine, bien que le mécanisme ne soit pas pleinement élucidé, le résidu Glu1 du domaine Te semble être indispensable à la cyclisation du peptide. Ce résidu n'a en revanche pas d'impact dans le cas d'une hydrolyse sans cyclisation.

Le domaine Te n'est pas le seul domaine impliqué dans la libération du NRP synthétisé. Il est par exemple substitué par un domaine réductase chez les NRPSs impliquées dans la synthèse de peptaibols, des NRPs porteurs d'un groupement alcool en C-terminale, qui sont notamment produits par *Trichoderma* spp. [56]. L'activité du domaine réductase permet la réduction de l'acyl-thioester en son alcool primaire, entraînant la libération du peptide avec un groupement alcool en C-terminale.

Chez les organismes fongiques, le domaine Te est très souvent absent et remplacé par un domaine de condensation terminal. On retrouve par exemple ce type de domaine au sein de la NRPS du ferrichrome chez *S. pombe* (**Figure 3C**) [23]. Bien qu'ils soient tous deux impliqués dans la libération et/ou cyclisation du peptide produit, le domaine thioestérase et le domaine de condensation terminal présentent une structure et un mécanisme d'action très différents. Dans le cas du domaine thioestérase, le peptide est d'abord fixé à la sérine du site actif avant d'être clivé. À l'inverse, il n'y a pas d'attachement du peptide au domaine de condensation terminal. L'histidine du site catalytique dé-protone directement le groupement amine du thio-

peptide porté par le domaine de thiolation précédent, induisant ainsi le mécanisme de cyclisation [57].

2.2.1.2. Les domaines secondaires

Les domaines secondaires sont des domaines qui ne sont pas indispensables à l'assemblage des monomères pour donner les peptides. Ces domaines interviennent en revanche dans la modification des monomères du peptide non-ribosomique au cours de la synthèse.

- **Domaine d'épimérisation**

Les domaines d'épimérisation présentent également des similitudes de structure avec les domaines de condensation, avec la présence notamment de la même histidine conservée au sein du site catalytique. Leur rôle consiste à convertir les acides aminés initialement reconnus et activés sous leur forme L, en leur énantiomère de forme D. La structure du domaine d'épimérisation présent au sein de la NRPS de la tyrocidine a été élucidée (**Figure 2C**) [58]. L'épimérisation intervient généralement directement sur le L-aminoacyl fixé au bras phosphopantéthéine, avant l'étape de condensation liant les deux monomères issus de modules successifs [59]. Néanmoins, la réaction d'épimérisation peut également être catalysée après la condensation, sur le complexe peptidyl-S-domaine T plutôt que sur le groupement L-aminoacyl-S-domaine T [60].

- **Domaine de méthylation**

La méthylation du monomère sélectionné au cours de l'assemblage du NRP est effectuée par un domaine de méthylation possédant une activité méthyltransférase permettant l'addition d'un groupement méthyle sur l'acyl thioester. Cette réaction peut s'effectuer avant ou après l'établissement de la liaison peptidique par le domaine de condensation [30]. Les domaines de méthylation ne sont généralement pas situés entre des domaines distincts, mais plutôt intégrés directement au sein d'un domaine d'adénylation. Deux positions sont souvent observées : entre les signatures A2 et A3 ou entre les signatures A8 et A9 [61].

- **Domaine de formylation**

La présence de domaines de formylation a été rapportée chez certaines synthétases comme par exemple la NRPS de la gramicidine linéaire [62]. Ce domaine catalyse la fixation d'un groupement formyle (-CH=O) sur le monomère sélectionné. Il se positionne généralement au sein du module d'initiation de la synthétase, juste avant le premier domaine d'adénylation (Module F-A-T-C). La cristallisation du module d'initiation de LgrA, NRPS de la gramicidine linéaire, a mis en évidence une rigidité de la zone de liaison entre le domaine de formylation et celui d'adénylation. Cette rigidité très peu habituelle au sein des NRPSs, qui sont des structures généralement très flexibles, témoigne d'une étroite interaction entre ces deux domaines [63].

- **Domaine de cyclisation**

Les domaines de cyclisation présentent une très forte similitude avec les domaines de condensation auxquels ils peuvent même se substituer. Ils sont impliqués dans la cyclisation de certains monomères au cours de la synthèse du NRP. Ils permettent notamment de générer les composés oxazoline et thiazoline en catalysant respectivement l'hétérocyclisation des résidus cystéine, ou sérine/thréonine [38,64].

- **Domaine d'oxydation**

Les domaines d'oxydation sont assez peu fréquents au sein des NRPSs. Leur implication dans la conversion de thiazoline et oxazoline en thiazole et oxazole, a été démontrée [26]. Ces domaines ont essentiellement été identifiés chez les synthétases de NRPS comportant ces structures, comme par exemple l'épothilone synthétase [65] ou la bléomycine synthétase [66]. La position des domaines d'oxydation au sein de la synthétase est variable puisqu'ils peuvent se positionner entre le domaine A et le domaine T comme dans le cas de la NRPS de bléomycine, ou entre le domaine T et le domaine C comme chez la NRPS de l'épothilone [67].

2.2.2. Les synthétases : chaînes d'assemblage de peptides non-ribosomiques

Le fonctionnement des NRPSs s'apparente à celui d'une chaîne d'assemblage de peptides où chaque domaine joue un rôle. Les monomères constituant le peptide non-ribosomique sont dans un premier temps reconnus spécifiquement par les domaines d'adénylation, puis fixés sur le bras phosphopantéthéique porté par les domaines de thiolation. Les domaines de condensation catalysent la liaison peptidique entre les monomères reconnus par 2 modules successifs. Les domaines secondaires sont responsables de la modification en cours de synthèse des monomères incorporés dans le peptide. Ces modifications peuvent avoir lieu avant ou après la réaction de condensation. L'activité du domaine final, que ce soit un domaine thioestérase, réductase, ou de condensation terminal, permet la libération du peptide synthétisé, parfois accompagnée d'une cyclisation.

Le mode d'assemblage linéaire, comme chez la tyrocidine synthétase (**Figure 2C**), est celui le plus fréquemment observé. Dans ce cas de figure, le nombre et l'ordre des monomères incorporés dans le peptide sont directement corrélés au nombre et à l'ordre des modules présents dans la synthétase. La NRPS débute alors par un module d'initiation (A-T) qui charge le premier monomère, puis se poursuit par autant de modules (C-A-T) que nécessaires pour compléter le NRP, avant de s'achever par un domaine terminal (Te, Re, ou C_{term}) permettant la libération du peptide produit [55]. Chez les NRPSs impliquées dans la biosynthèse des lipopeptides (peptides non-ribosomiques comportant une chaîne lipidique), le module d'initiation débute souvent par un domaine de condensation C_{start} établissant la liaison entre la partie lipidique et le premier monomère chargé. Cette conformation est par exemple observée chez la synthétase de la surfactine produite par *Bacillus* spp. [68].

Il existe d'autres modes de synthèse plus atypiques comme celui de la synthèse itérative. Ce mode de synthèse implique qu'un ou plusieurs modules de la NRPS accomplissent plusieurs cycles permettant d'incorporer à de multiples reprises un

même monomère (activé par un même domaine A) au sein du NRP. Le nombre de monomères incorporés dans le NRP n'est alors pas corrélé avec le nombre de modules présents au sein de la NRPS. C'est le cas par exemple de la gramicidine synthétase dont les deux sous-unités GrsA (activant D-Phe) et GrsB (activant Pro-Val-Orn-Leu) doivent effectuer deux cycles de synthèse pour assembler un décapeptide de 2 fois 5 monomères [69]. Le terme de synthèse itérative n'est pas uniquement employé lorsque l'itération d'un module complet de NRPS est nécessaire pour l'assemblage du NRP final, mais également lorsque l'un des domaines au sein de la NRPS fonctionne de manière itérative. C'est le cas par exemple lors de l'assemblage du ferrichrome chez *S. pombe* (**Figure 3C**) [22]. [23]. La synthétase du ferrichrome s'organise en 2 parties avec une structuration double A-T-C-T-C-T-C. Chacun des 2 domaines d'adénylation fonctionnels charge de manière itérative le même monomère sur 3 domaines de thiolation distincts. Les domaines de condensation établissent ensuite la liaison peptidique entre les monomères générant ainsi un cyclohexapeptide.

Une grande diversité des modes de fonctionnement des NRPSs a pu être mise en évidence, si bien qu'il est difficile d'établir des règles applicables pour l'ensemble des NRPSs. La synthétase de l'amonabactine, un sidérophore produit par *Aeromonas* spp., en est le parfait exemple [32]. Cette NRPS présente un fonctionnement itératif, alternatif et optionnel. Le caractère alternatif est porté par le domaine amoG activant soit Trp ou Phe, auquel viennent se fixer 2 dipeptides (Dhb-Lys) générés de manière itérative par les domaines amoE et amoF (2 cycles de synthèse). Enfin, de manière optionnelle dans deux cas sur quatre, un domaine amoH inclut une Gly entre le Dhb et la Lys du dipeptide décrit précédemment. En combinant les diverses possibilités, 4 peptides différents peuvent ainsi être assemblés par cette NRPS.

2.3. Diversité de structures et d'activités des NRPs

Les synthétases de peptides non-ribosomiques sont capables de générer des peptides présentant une formidable diversité de structure. L'analyse des peptides non-ribosomiques répertoriés dans Norine (<https://bioinfo.lifl.fr/norine/>), la seule base de données dédiée à l'étude des NRPs, permet de visualiser l'éventail de structures existantes (**Figure 7A**) [70].

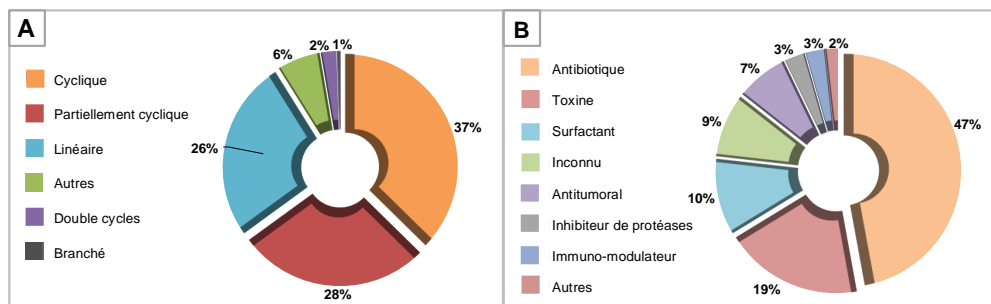


Figure 7 : Analyse des propriétés des peptides d'origine non-ribosomique répertoriés dans la base de données Norine en termes de structures (**A**) et d'activités (**B**). Certains NRPs présentant des activités multiples ont été pris en compte dans plusieurs des catégories représentées sur le diagramme B.

Parmi les 1191 peptides répertoriés à ce jour, les 3 formes de NRPs prédominantes sont les structures cycliques (37 %), partiellement cycliques (28 %), et linéaires (26 %). Il existe également des peptides non-ribosomiques présentant une structure branchée comme par exemple la famille des pyoverdines produites par *Pseudomonas* spp. [71]. Une grande variabilité est également à souligner concernant la taille des peptides non-ribosomiques qui peuvent être constitués de 2 à 26 monomères. La moyenne du nombre de monomères incorporés dans les NRPs répertoriés est à 9,6, et les tailles les plus représentées sont 7 et 8 monomères (observés pour 165 et 162 NRPs respectivement).

Outre la pluralité de structure, la diversité se manifeste également par la présence dans les NRPs de monomères divers tels que des lipides, des polykétides, des sucres, et bien entendu des acides aminés. Au total, plus de 500 monomères différents ont été identifiés comme des constituants de NRPs [72]. Parmi ces monomères, certains sont incorporés directement en tant que tels, et d'autres sont le résultat de modifications générées au cours de l'assemblage du peptide grâce à l'activité des domaines de la NRPS. Le grand nombre de modifications des monomères (épimérisation, méthylation, oxydation, etc.) catalysées par les NRPSs accroît la diversité des peptides assemblés. En matière de potentiel de diversité, la voie non-ribosomique se place bien au-dessus de la voie ribosomique où seuls 20 acides aminés (dits protéinogènes) sont utilisés pour l'assemblage des protéines [26].

La très grande variété structurale des peptides non-ribosomiques est la source même de la diversité de leurs activités biologiques et de leurs propriétés physico-chimiques. La pluralité des applications possibles est parfaitement illustrée par l'analyse des

activités des NRPs répertoriés au sein de la base de données Norine (**Figure 7B**). L'activité antibiotique (antibactérienne / antifongique) est celle qui est la plus fréquemment observée, puisqu'elle concerne 47 % des NRPs répertoriés. Parmi les principaux NRPs antibiotiques, on peut citer la tyrocidine (dont la voie de biosynthèse est décrite sur la **Figure 2**) et la gramicidine, produites par *B. brevis*, qui présentent notamment une activité antibactérienne contre les organismes de type gram positifs [73]. La pénicilline, produite par *P. chrysogenum* et mise en évidence par Alexander Fleming, est obtenue à partir d'un peptide non-ribosomique [74]. La vancomycine produite notamment par *Amycolatopsis orientalis* et par de nombreux autres actinomycètes, présente une activité antibactérienne contre *Staphylococcus aureus*, un pathogène très répandu dans les produits alimentaires [75]. Enfin, une activité antifongique a été démontrée chez les peptaibols produits par *Trichoderma* spp. [76] et les echinocandines isolés chez *Aspergillus* spp. [77].

L'activité cytotoxique concerne 19 % des NRPs répertoriés dans la base de données. On retrouve notamment des toxines impliquées dans des cas d'intoxications humaines comme par exemple les microcystines produites par plusieurs cyanobactéries [78]. La biosynthèse de composés phytotoxiques ou insecticides a également été décrite chez plusieurs champignons pathogènes comme par exemple la production de destruxine et tentoxines par *Alternaria brassicae* (pathogène répandu du colza) [79,80], d'enniatiines par *Fusarium* spp. (phytopathogène de nombreuses cultures) [81], ou de beauvericines par *Beauveria bassiana* [82].

Les propriétés surfactantes qui sont caractéristiques de 10 % des NRPs concernent en tout premier lieu les lipopeptides, comportant une chaîne d'acide gras. La famille des surfactines produites par *Bacillus* spp. constitue le meilleur exemple de biosurfactant avec des propriétés moussantes et émulsifiantes remarquables. Les autres familles de lipopeptides synthétisées par *Bacillus* spp., les iturines et les fengycines, présentent également une activité surfactante [83].

Parmi les autres activités fréquemment observées, on peut mentionner les propriétés immuno-modulatrices et anti-inflammatoires retrouvées par exemple chez les cyclosporines produites par *Beauveria nivea* [84], l'activité inhibitrice de protéases des anabaenopeptines synthétisées par *Microcystis aeruginosa* [85], ou les propriétés sidérophores de nombreux NRPs comme le ferrichrome produit par *S. pombe* (décrit **Figure 3**) [23,64]. Enfin, certains NRPs comme la tryprostatine B (dérivée de brevianamide F) ou les astines sont décrits comme des composés à activité anti-tumorale [86,87].

3. Les astines : pentapeptides cycliques non-ribosomiques

3.1. Généralités et structures

Les astines sont des pentapeptides cycliques non-ribosomiques, qui ont en premier lieu été isolés à partir des racines de la plante *Aster tataricus* utilisée traditionnellement en médecine chinoise et japonaise [88]. Divers terpénoïdes et saponines ont également été détectés au sein de cette plante [89]. La présence combinée de ces composés et des pentapeptides cycliques confère des propriétés thérapeutiques intéressantes à *A. tataricus* avec notamment des effets diurétique, antitussif, anti-ulcère, antiviral, antibactérien et anti-tumoral [90,91].

À ce jour, 16 formes d'astine différentes ont été isolées à partir d'*A. tataricus*, puis caractérisées via des techniques de spectrométries et de RMN [92] (**Figure 8**).

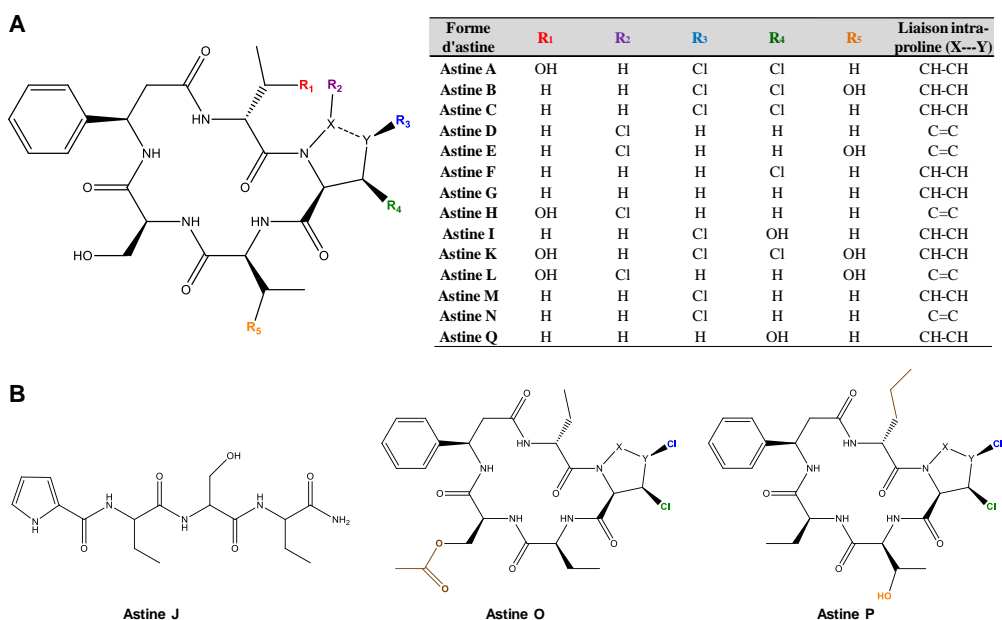


Figure 8 : Structure des différentes formes d'astine isolées à partir des racines d'*Aster tataricus*. **A.** Organisation générale des astines et éléments de différenciation entre les différentes formes : R₁, R₂, R₃, R₄, R₅, et liaison X---Y. **B.** Structure des formes d'astine J (linéaire), O, et P présentant certaines variations vis-à-vis du schéma décrit sur le panneau A (mis en évidence par la couleur brune).

Dans un premier temps, seules les formes A, B et C ont été identifiées et étudiées par les équipes de Morita *et al.* [86,88,93]. Les mêmes formes d'astine ont parfois été mentionnées dans la littérature en tant qu'asterinin [94] ou asterin [95]. Par la suite, 13 nouvelles formes d'astine ont été successivement isolées et caractérisées [96–98], ainsi que 2 formes dérivées d'astine nommées tataricine A et B [99]. Les différentes formes d'astine présentent une structure cyclique (à l'exception de l'astine J linéaire),

et comportent 5 acides aminés dont 2 sont protéinogènes (sérine, proline) et 3 non-protéinogènes (acide 2-amino-butérique, allo-thréonine, β -phénylalanine). Le résidu proline est porteur d'aucun, un, ou deux groupements chlore (Cl) en fonction de la forme d'astine considérée [91].

Récemment, la production de plusieurs formes d'astine a été mise en évidence chez un champignon filamenteux, *Cyanoderrella asteris*, contaminant récurrent des cultures d'*A. tataricus* (**Figure 9**). La présence de ce champignon a été mise en évidence essentiellement au niveau des racines d'*A. tataricus*, mais également au sein des feuilles et des inflorescences de la plante. *Cyanoderrella asteris* semble coloniser la plante en tant qu'endophyte et ne présente en apparence aucun comportement saprophytique [100,101].

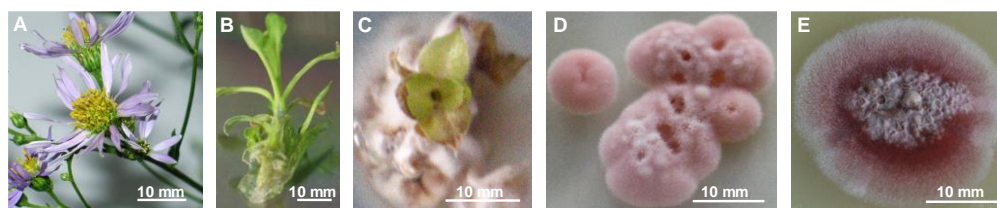


Figure 9 : Illustrations relatives à l'isolement de l'endophyte *Cyanoderrella asteris* à partir de la plante *Aster tataricus*. Les différentes illustrations ont été adaptées à partir des figures provenant de l'étude de Jahn *et al.* [100]. **A.** Observation visuelle de l'inflorescence d'*A. tataricus*. **B.** Plantule d'*A. tataricus* stérile se développant sur un milieu contenant des hormones de croissance. **C.** Aspect de la plantule après 1 mois de culture, et premières étapes du développement du champignon endophyte. **D.** Développement précoce de *C. asteris* isolé sur un milieu agar pomme de terre (PDA). **E.** Aspect de *C. asteris* après 55 jours de culture sur milieu PDA.

Ce champignon présente 95 % d'homologie de séquence génomique avec le champignon *Talaromyces islandicus* (anciennement nommé *Penicillium islandicum*) producteur de cyclochlorotine [102]. La cyclochlorotine, aux propriétés cytotoxiques, présente une très grande similarité de structure avec l'astine B puisqu'elle ne varie qu'au niveau d'un seul monomère [90]. La proximité génétique de *C. asteris* avec un champignon produisant un pentapeptide cyclique, est un élément supplémentaire permettant de conjecturer que ce champignon est le producteur naturel d'astine plutôt qu'*A. tataricus*. De plus, à ce jour, la présence d'une voie de synthèse non-ribosomique n'a pu être mise en évidence chez aucun eucaryote appartenant au clade taxonomique plantae.

L'analyse des profils de production d'astine par *C. asteris* a mis en évidence son incapacité à produire l'ensemble des formes d'astine qui ont été isolées à partir des racines d'*A. tataricus* [101]. Seules les formes C, F et G ont pu être détectées en cultivant l'endophyte en milieu à base de dextrose et pomme de terre (PDB) ou d'extrait de malt (MEB) (**Figure 10A**). La forme C de l'astine, dont la structure est présentée **Figure 10B**, est celle qui est très majoritairement produite par *C. asteris*. Le fait qu'*A. tataricus* soit capable de prélever les composés produits par son

endophyte, puis de les modifier par la suite, constitue l'hypothèse la plus probable pour expliquer la présence de formes supplémentaires d'astine chez la plante [101].

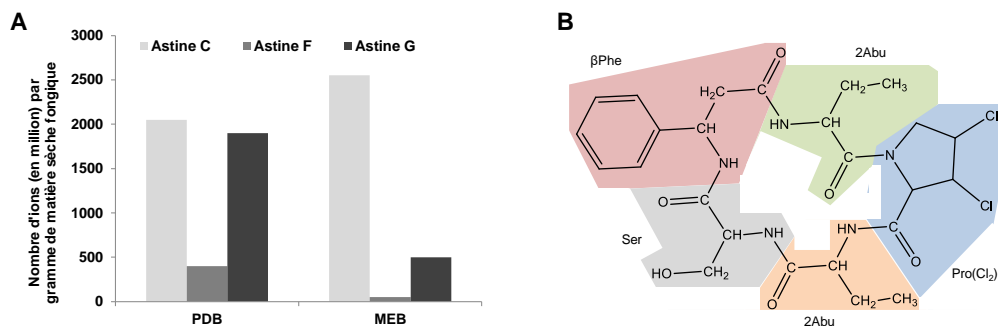


Figure 10: **A.** Profil de production des différentes formes d'astine produites par *Cyanodermea asteris* en culture pendant 28 jours en milieu PDB et MEB. **B.** Structure moléculaire de l'astine C, principale forme produite par *C. asteris*, comprenant 5 monomères : deux L-2-aminobutyrate (2Abu), une L-β-phénylalanine (βPhe), une L-cis-3,4-dichloroproline [Pro(Cl)₂] et une L-serine (Ser).

3.2. Activités

Les premières formes d'astine isolées, les astine A, B, et C, ont été les plus étudiées en termes d'activités. Dans un premier temps leur potentiel anti-tumoral a été mis en évidence *in vitro* contre des lignées de cellules cancéreuses murines de type sarcome S-180 [86,88]. Par la suite, la cytotoxicité des astines B et C a été démontrée contre des lignées cellulaires de cancer gastrique humain BGC-823 avec une IC₅₀ (concentration pour laquelle la moitié des cellules meurent) respectivement de 19,2 mg/mL et 3,3 mg/mL [98]. En ce qui concerne l'astine B, il a été démontré qu'une thionation (thioastine B) pouvait accroître son activité anticancéreuse [103]. En revanche, les formes d'astine présentant aucun ou un seul groupement chlore sur le résidu proline (Astine D-J) ne possèdent pas de propriétés anti-tumorales [97]. Par opposition, le résidu cis-3,4-di-Cl proline est considéré comme étant de première importance en matière d'activité anticancéreuse pour les astines A, B et C [91,97,104].

Afin d'étudier plus en détail les mécanismes d'action des composés de type astine, plusieurs études se sont intéressées à l'activité de composés synthétiques analogues des astines [105–107]. Les cyclo-pentapeptides synthétisés différaient des astines par la présence des résidus non-protéinogènes Aib (acide α-amino-iso-butyrique) et/ou (S)β³-hPhe, ainsi que celle d'un groupement SO₂-NH substituant une des liaisons peptidiques. Ces composés synthétisés ont été testés contre des lignées cellulaires cancéreuses humaines de thyroïde (NPA). L'activité la plus intense a été obtenue pour le pentapeptide synthétisé présentant la plus grande flexibilité de structure (conférée notamment par la présence du groupement SO₂) [105,106]. L'étude menée par la suite par Cozzolino *et al.* (2005) a permis d'élucider partiellement le mode d'action des composés de type astine. Cette étude a montré que le cyclopeptide synthétique (Pro –

Abu – Ser - (S) β^3 -hPhe - CH₂-SO₂-NH-Abu) agit comme un activateur des caspases 3, 8 et 9, induisant par la suite l'apoptose des cellules tumorales (NPA) [107].

L'équipe de Wang *et al.* a réalisé une étude plus approfondie concernant les activités de l'astine B [90]. Cette forme d'astine a montré une activité cytotoxique sur des cellules hépatiques humaines (L-02), avec une diminution de la viabilité cellulaire de 39,22 % à 60 μ M après 48 h. Le traitement à l'astine B induit la mort des cellules par apoptose. Celle-ci se matérialise par une augmentation de l'expression du composé pro-apoptotique Bax, ainsi qu'une diminution de l'expression du composé anti-apoptotique Bcl-2. Le ratio Bax / Bcl-2 s'en trouve considérablement augmenté, favorisant l'apoptose des cellules (**Figure 11A**).

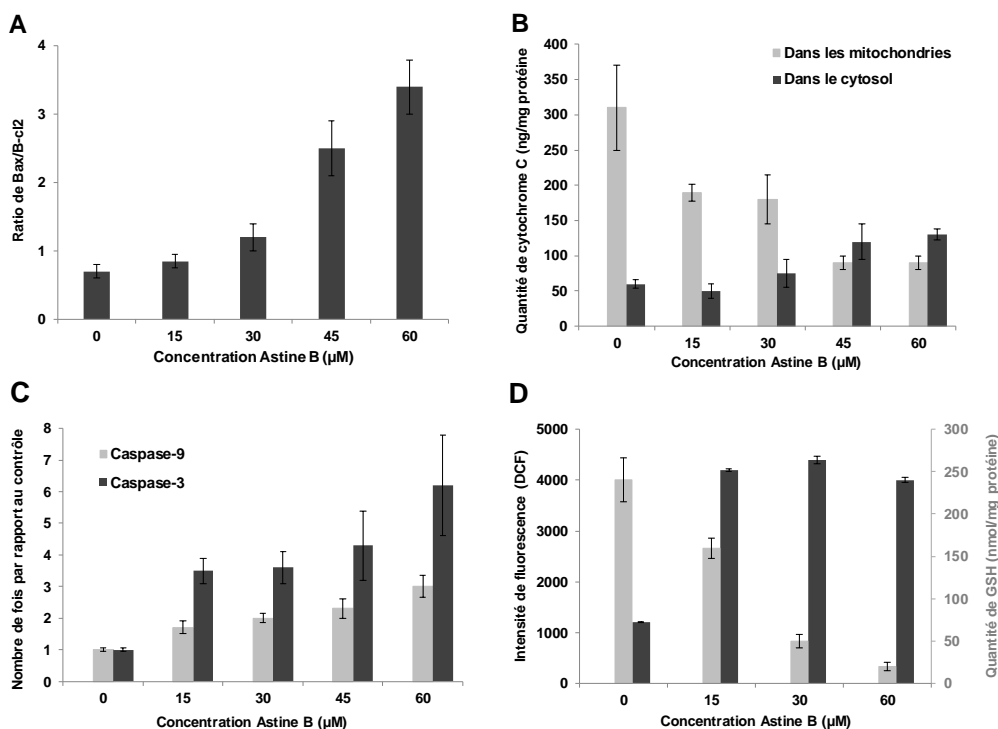


Figure 11 : Différents effets observés *in vitro* chez des cellules hépatiques humaines L-02 après traitement avec des concentrations variables d'astine B. Figure adaptée à partir des résultats obtenus par Wang *et al.* (2014) [90]. **A.** Ratio de Bax / Bcl-2 mesuré par densitométrie après 24 h de traitement. **B.** Quantité de cytochrome C présent au sein des mitochondries et dans le cytosol, mesurée via un kit ELISA après 24 h de traitement. **C.** Activités enzymatiques des caspases-3 et -9 mesurées après 24 h de traitement. **D.** Mesure de la présence de composés induisant un stress oxydatif ou au contraire à caractère antioxydant. Le taux de dérivés réactifs de l'oxygène (en gris foncé) a été mesuré via un composé spécifique fluorescent (DCFH-DA) après 12 h de traitement. Le taux intracellulaire du composé antioxydant, le glutathion (en gris clair), a été déterminé après 24 h de traitement.

Les composés pro-apoptotiques ont un effet perméabilisant sur la membrane des mitochondries, ce qui explique la libération dans le cytosol du cytochrome C contenu dans les mitochondries (**Figure 11B**). Le cytochrome C est un activateur des caspases-9, qui activent par la suite les caspases-3, induisant l'apoptose des cellules. Une augmentation des activités des caspases-3 et -9 a bien été démontrée à la suite du traitement à l'astine B (**Figure 11C**). Par ailleurs, le traitement à l'astine B entraîne une augmentation du taux d'espèces réactives de l'oxygène, associée à une baisse du taux de glutathion (antioxydant), entraînant un stress oxydatif favorisant l'apoptose des cellules L-02 (**Figure 11D**). En somme, le traitement à l'astine B entraîne à la fois une inflammation due à un stress oxydatif, ainsi qu'une apoptose des cellules induite via un dysfonctionnement mitochondrial entraînant la libération du cytochrome C activant les caspases-3 et -9 [90]. En parallèle, des essais *in vivo* réalisés chez des souris ont montré que, malgré son potentiel antinéoplasique, l'astine B pouvait également présenter une certaine toxicité et entraîner notamment l'apparition de lésions hépatiques [90].

L'astine C a également été évaluée en tant que molécule immuno-modulatrice vis-à-vis de cellules ganglionnaires (lymphocyte T) de souris (C57BL/6), préalablement activées ou pas par la concavaline A [108]. L'astine C présente une forte activité inhibitrice de la prolifération des lymphocytes T activés, et une activité faible vis-à-vis des cellules inactivées (**Figure 12A**).

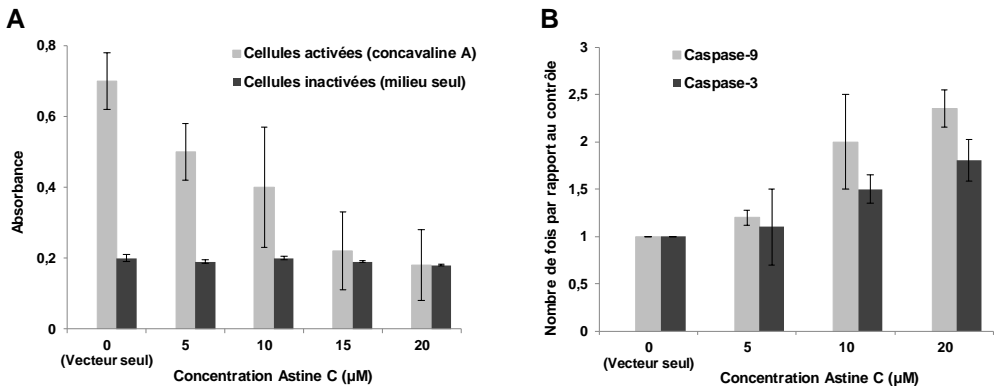


Figure 12 : Effets observés chez les cellules ganglionnaires (lymphocyte T) de souris soumises à un traitement de concentrations variables d'astine C pendant 24 h. Figure adaptée à partir des résultats obtenus par Shen *et al.* (2011) [108]. **A.** Mesure via la méthode MTT de la prolifération des cellules T actives ou inactives après traitement à l'astine C. **B.** Activités enzymatiques des caspases-3 et -9 mesurées après traitement à l'astine C.

En complément de l'inhibition de la prolifération, un phénomène d'apoptose a également été observé chez les cellules activées après traitement à l'astine C. De la même manière que dans le cas de l'astine B, l'astine C induit une augmentation de l'expression du composé pro-apoptotique Bax, entraînant une libération du cytochrome C initialement contenu dans les mitochondries. La présence du cytochrome C dans le cytosol permet ensuite une augmentation de l'activité des

caspases-3 et -9 (**Figure 12B**). En revanche, aucun facteur laissant supposer une induction de l'apoptose via un stress du réticulum endoplasmique n'a pu être mis en évidence. Aux vues de ses effets sur les lymphocytes T activés, des essais *in vivo* ont été menés sur des souris afin d'évaluer l'impact d'un traitement à l'astine C sur la colite (inflammation de l'intestin). Le traitement à l'astine C a montré des effets bénéfiques, manifestés notamment par une limitation de la perte de poids et une diminution significative du ratio poids / longueur de l'intestin chez les sujets traités. L'augmentation du ratio poids / longueur est un marqueur de l'inflammation de l'intestin. L'observation visuelle des tissus a montré une nette diminution de l'inflammation et de la nécrose de l'épithélium chez les sujets traités. Cette étude démontre le potentiel anti-inflammatoire de l'astine C, qui s'ajoute à ses propriétés antinéoplasiques décrites précédemment [86]. La capacité de l'astine C à agir uniquement sur les cellules T activées, et pas sur les cellules inactives, la positionne comme un composé prometteur pour répondre aux pathologies inflammatoires.

3.3. Modes de production

Plusieurs méthodes ont été développées afin de produire par synthèse chimique les formes d'astines naturelles [92] ou des dérivés synthétiques [105–107]. Les coûts engendrés par l'utilisation de certains précurseurs, notamment ceux relatifs au résidu proline chloré, limitent la production d'astines par cette approche à une échelle de laboratoire. Hormis la synthèse chimique, il existe deux moyens d'obtenir ces pentapeptides cycliques : soit via une extraction à partir des racines d'*A. tataricus*, soit directement à partir d'une culture du champignon endophyte *C. asteris*.

3.3.1. Extraction à partir d'*Aster tataricus*

L'extraction à partir d'*A. tataricus* nécessite une grande quantité de racines sèches. Les racines, éventuellement broyées préalablement dans de l'azote liquide [101], sont plongées en immersion dans du méthanol (MeOH) à température ambiante [95]. L'extrait de méthanol obtenu est ensuite traité avec du n-butanol, après un prétraitement au chlorure de méthylène [86]. Les astines chlorées sont ensuite purifiées à partir de cette fraction n-butanol via plusieurs étapes de chromatographie. Une méthode possible consiste à réaliser une première élution en phase normale avec CHCl₃-MeOH (9:1), puis une seconde en phase inverse (RP-8 à 60 % MeOH) [95]. Les rendements obtenus en utilisant cette approche sont très faibles. En ce qui concerne l'astine C un rendement de 8 g pour 50 kg de racines a été observé, ce qui correspond à 0,16 mg/g de matériel végétal [108].

3.3.2. Production via l'endophyte *Cyanodermella asteris*

Récemment, l'isolement du champignon endophyte *C. asteris*, à partir des racines d'*A. tataricus*, a ouvert de nouvelles perspectives, notamment en termes de production d'astine C. Ce champignon se développe sur des milieux à base d'extraits de pomme de terre (PDA, PDB) ou enrichi en extrait de malt (MEA, MEB). Néanmoins, sa croissance est très lente puisque son développement en milieu solide ou liquide

nécessite au minimum 4 semaines [101]. Ce faible taux de croissance constitue le principal frein à la production d'astine. Néanmoins, ce champignon sécrète directement les astines produites dans le milieu de culture liquide, ce qui facilite l'extraction et la purification des composés [101]. Une fois le champignon éliminé par centrifugation, les astines peuvent être directement extraites du milieu liquide par un traitement au n-butanol, puis éventuellement purifiées par chromatographie comme décrit précédemment. Cette approche a permis l'extraction de l'astine C, à partir d'une culture de *C. asteris* âgée de 3 semaines, avec un rendement après purification de 21 mg/L [16].

4. Stratégies d'optimisation de la production homologue de NRPs par un champignon filamenteux

Le développement des technologies de séquençage et d'analyse des génomes a permis de mettre en évidence l'immense potentiel de production de métabolites secondaires des organismes fongiques [109,110]. Ce potentiel visible à l'échelle du génome ne se matérialise souvent pas dans les profils de production obtenus en fermentant la souche. L'absence ou la faible production d'un composé peut être due au caractère cryptique (non-exprimé) des gènes impliqués dans sa biosynthèse, et / ou à des conditions de cultures inadaptées à sa production [111]. Deux approches sont possibles pour développer la production homologue d'un composé d'intérêt : (1) en modifiant la souche productrice à l'échelle génomique aléatoirement ou par ingénierie génétique, (2) en optimisant les conditions de culture de manière à favoriser sa production.

4.1. Optimisation génétique de la souche

L'optimisation génétique d'une souche pour obtenir un mutant présentant de hauts taux de production est une approche qui était déjà utilisée bien avant l'ère post-génomique. L'utilisation d'agents mutagènes chimiques, de rayons X, et d'ultraviolet, a par exemple permis de générer des mutants d'*Aspergillus niger* présentant un taux de production élevé en acide citrique [112].

L'avènement de l'ère post-génomique a ouvert de nouvelles perspectives en termes d'optimisation génétique des souches fongiques. Une approche possible consiste à insérer plusieurs copies du gène ou du cluster de gènes impliqué dans la biosynthèse du composé d'intérêt. La plupart des souches de *P. chrysogenum* surproductrices de pénicilline (utilisées à l'échelle industrielle) comportent plusieurs copies du gène codant pour la synthétase d'ACV (précurseur de la pénicilline) [113]. Une autre stratégie couramment utilisée consiste à remplacer le promoteur régulant l'expression du gène impliqué dans la biosynthèse du composé d'intérêt par un promoteur capable d'induire une transcription plus importante. Chez *Aspergillus nidulans*, le remplacement des promoteurs des gènes NRPS *ivoA*, *ivoB*, et *ivoC*, par le promoteur fort P_{alcA} a permis d'accroître l'expression de ces gènes et d'augmenter les taux de production des NRPs (pigments de couleur brune) [114]. Chez *Fusarium fujikuroi*, la même stratégie a été utilisée, cette fois-ci en utilisant le promoteur Tet-On induit en présence de tétracycline, afin d'activer la production de la PKS-NRPS de l'equisetine dont le gène était initialement cryptique [115]. Alternativement, l'expression d'un gène codant pour un composé d'intérêt peut également être accrue en sur-exprimant des facteurs de transcription capables de réguler positivement sa transcription. De nouveau chez *F. fujikuroi*, la surexpression d'un facteur de transcription spécifique a permis d'obtenir la production de l'apicidine F, un peptide non-ribosomique qui n'avait jamais été détecté avant cette manipulation génétique [116].

Chez les champignons filamenteux, les clusters de gènes impliqués dans la biosynthèse de métabolites secondaires sont souvent localisés au niveau des télomères

des chromosomes. Cette région chromosomique adopte généralement une conformation sous forme d'hétérochromatine où les niveaux d'expression des gènes sont très faibles [117,118]. L'absence ou la faible production d'un NRP d'intérêt peut s'expliquer par le positionnement du cluster de gènes correspondant au sein d'une de ces zones où les gènes sont peu ou pas exprimés. Il est possible de lever cette limitation en influant sur les mécanismes de structuration de la chromatine afin notamment de limiter la mise en place de l'hétérochromatine [119]. La formation de l'hétérochromatine est médiée en partie par la méthylation des histones H3K9, H3K27 et H4K20 [118]. Westphal *et al.* (2019) ont développé une stratégie permettant de limiter la formation de l'hétérochromatine par le biais de la délétion du gène *ktm6*, codant pour une méthyltransférase impliquée dans la méthylation de l'histone H3K27. La délétion de ce gène a permis d'augmenter les rendements de production homologue d'un NRP, la fusaotaxin A, chez *Fusarium graminearum* [120]. À l'inverse, la méthyltransférase codée par le gène *laeA*, très conservée chez l'ensemble des champignons filamenteux, permet de limiter la formation de l'hétérochromatine en neutralisant la méthylation de l'histone H3K9. La surexpression du gène *laeA* est l'une des stratégies fréquemment employées pour augmenter les niveaux d'expression des clusters de gènes impliqués dans la biosynthèse de métabolites secondaires chez les champignons filamenteux [118,121,122]. Il a été démontré que le niveau de régulation de l'expression des gènes par la protéine LaeA dépendait de la localisation du gène ou du cluster de gène au sein du chromosome. La translocation d'une partie ou de la totalité des gènes du cluster vers une zone chromosomique plus favorable constitue donc également une approche possible pour augmenter les niveaux d'expression de ces gènes et potentiellement accroître les rendements de production du métabolite secondaire concerné [123].

Une autre approche pour améliorer la production d'un composé d'intérêt consiste à inactiver d'autres voies de biosynthèse qui peuvent concurrencer la production, notamment en monopolisant certains précurseurs communs. Par exemple, chez *G. lozoyensis*, les taux de production de la pneumocandine ont été doublés après la délétion du gène impliqué dans la biosynthèse de la mélanine [124]. La même stratégie a été appliquée chez *A. nidulans* avec la délétion des clusters de gènes impliqués dans la biosynthèse de l'emericellamide et de la sterigmatocystine, deux métabolites majoritairement produits, entravant la synthèse d'autres composés [125].

Récemment, le développement de la technologie CRISPR-cas9 a ouvert de nouvelles perspectives en termes d'édition des génomes. De nombreux outils relatifs à cette technique sont déjà disponibles pour faciliter l'édition des génomes de champignons filamenteux, comme par exemple chez *A. niger* [126].

Cependant, l'absence d'outils moléculaires spécifiques pour un grand nombre d'organismes fongiques, et notamment ceux nouvellement identifiés, constitue bien souvent un facteur limitant à l'utilisation de l'approche génétique pour l'optimisation de la production d'un composé d'intérêt.

4.2. Optimisation des conditions de culture

Cette seconde approche, basée sur l'optimisation des conditions de culture de la souche, est applicable à tout organisme fongique cultivable. Elle consiste à définir les conditions de culture les plus adaptées pour favoriser la production du composé d'intérêt.

L'approche la plus évidente consiste très souvent à optimiser la composition du milieu de culture afin de trouver celle la plus appropriée pour la croissance de la souche et sa productivité [127]. L'optimisation du milieu de culture, avec notamment une augmentation des proportions de D-mannitol et KH_2PO_4 , a permis de multiplier par 6 la productivité en pneumocandine d'un même mutant de *Zalerion arboricola* [128]. Les différents paramètres de cultures (température, pH, agitation, méthode d'inoculation, etc.) sont également à prendre en considération en termes d'optimisation.

L'ajout dans le milieu de certains précurseurs nécessaires à la biosynthèse du composé d'intérêt est un autre moyen d'augmenter les taux de production. L'ajout d'acide 5-hydroxytetradecanoïque, précurseur de la verlameline, dans le milieu de culture de *Verticillium lamellicola* a par exemple permis d'augmenter la production de ce NRP de manière proportionnelle à la quantité de précurseur ajouté [129]. Cette approche est déjà couramment utilisée chez des organismes bactériens pour augmenter la production de certains lipopeptides comme la pseudofactine chez *Pseudomonas fluorescens* [130], ou la fengycine chez *Bacillus subtilis* [131]. À l'inverse la production de certains métabolites peut être stimulée dans des conditions de carence en certains éléments. La synthèse de xanthocilline et d'isocyanide chez *Aspergillus fumigatus* est par exemple positivement impactée en situation de carence en cuivre [132].

Enfin, le développement d'une méthode de culture alternative peut induire une modification du métabolisme de la souche considérée et impacter positivement sa productivité. Chez les organismes fongiques, l'utilisation d'un système de culture en cellules immobilisées a permis à de nombreuses reprises d'augmenter les rendements de production. Par exemple, la production de plusieurs métabolites (acide citrique, acide gluconique, glycosyl hydrolase, etc.) a pu être augmentée en cultivant *A. niger* de manière immobilisée [133–135]. Lorsque les cellules sont immobilisées sur un support, la récupération et la purification des métabolites sécrétés dans le milieu sont facilitées puisque l'étape de séparation cellules/milieu n'est plus nécessaire. Ce procédé constitue donc une approche prometteuse pour optimiser la production homologue d'un composé naturellement sécrété dans le milieu de culture.

5. Production hétérologue de peptides non-ribosomiques au sein d'usines cellulaires fongiques : de l'exploitation du génome à la production optimisée (Article)

- **Mise en contexte**

Comme dans le cas de *Cyanoderma asteris* et de la production d'astine, de nombreux organismes présentent un taux de croissance et de production de peptides non-ribosomiques faible ou inexistant. Cela peut s'expliquer par l'existence de clusters de gènes cryptiques (non exprimés) ou par des conditions de culture en laboratoire inadaptées. L'ingénierie génétique permet dans de nombreux cas d'augmenter les taux de production chez la souche homologue (naturellement productrice), mais elle ne peut être appliquée que si des outils moléculaires sont disponibles pour la souche considérée. Dans le cas contraire, le transfert de la voie de biosynthèse du peptide non-ribosomique d'intérêt au sein d'un hôte hétérologue, dont les conditions de cultures sont plus favorables, est bien souvent la solution adoptée. Cette revue de la littérature présente les différents outils bioinformatiques, ainsi que les étapes à considérer pour mettre en place une production hétérologue de peptides non-ribosomiques.

- **Résumé**

Les organismes fongiques sont réputés pour être des producteurs prolifiques de métabolites secondaires, et notamment de peptides non-ribosomiques (NRPs). La complexité structurale de ces composés leur confère des activités biologiques intéressantes, telles que des propriétés antibiotiques, anticancéreuses et anti-inflammatoires. Les progrès récents en matière de séquençage ont permis de générer une quantité incroyable de données génomiques fongiques provenant de centaines d'espèces microbiennes. Ces données ont révélé la présence d'un nombre inattendu de clusters de gènes potentiellement impliqués dans la synthèse de NRPs. Cela positionne les espèces fongiques comme une mine d'or en matière de découverte de nouveaux composés possiblement bioactifs. Étant donné que le taux de production des NRPs chez les producteurs natifs est souvent nul ou faible en raison de gènes silencieux ou de conditions de culture non adaptées, le transfert de leur voie de biosynthèse chez des hôtes hétérologues est souvent nécessaire. Cette revue permet de traiter le sujet de la découverte des NRPs dans son ensemble, de l'identification des clusters de gènes jusqu'à la production hétérologue dans des hôtes fongiques. Les principaux outils et méthodes pour l'exploitation des génomes seront présentés en mettant l'accent sur les particularités des systèmes fongiques par rapport aux systèmes bactériens. Les différentes étapes de la reconstitution de la voie de biosynthèse des NRPs au sein d'usines cellulaires fongiques hétérologues seront présentées, ainsi que les facteurs clés à prendre en compte pour maximiser la productivité. Plusieurs exemples seront développés afin d'illustrer le potentiel de la production hétérologue

à la fois en termes de découverte de nouveaux composés non caractérisés prédits *in silico*, et en termes d'augmentation de productivité pour des composés bioactifs naturels d'intérêt.

- **Mots Clés**

Outils bioinformatiques, Cellules fongiques, Exploitation du génome, Production hétérologue, Large cluster de gènes, Enzyme multi-modulaire, Peptide non-ribosomique, Synthétase de peptide non-ribosomique, Métabolite secondaire

- **Principales conclusions**

Le récent développement des technologies de séquençage à haut débit et des méthodes d'annotation ont permis de mettre en évidence le potentiel des organismes fongiques en matière de production de métabolites secondaires. Aujourd'hui, de nombreux outils bioinformatiques facilitent l'identification et la prédiction des structures des composés issus des voies non-ribosomiques. Néanmoins, les NRPs et NRPSs d'origine fongique restent sous-représentés au sein des bases de données par rapport à ceux d'origine bactérienne. Les outils de prédiction reposant essentiellement sur des mécanismes de comparaison, l'analyse des NRPs fongiques demeure donc imprécise dans la plupart des cas. Les composés identifiés *in silico* ne sont très souvent que peu ou pas produits chez le producteur natif, parce que le cluster de gènes correspondant est cryptique, ou que les conditions de cultures sont inappropriées. Plusieurs plateformes cellulaires fongiques ont été développées afin de mettre en place une production hétérologue de ces composés. Entre autres, *S. cerevisiae* et *Aspergillus* spp. ont été largement utilisés. En ce qui concerne la production hétérologue des NRPs, plusieurs étapes peuvent se révéler limitantes. Le transfert des séquences NRPS de longue taille (jusqu'à plusieurs dizaines de kilobases) constitue un premier obstacle. Plusieurs méthodologies, telles que le TAR cloning, ExRec, l'assemblage Gibson, et le clonage USER, ont été développées pour faciliter la reconstruction des larges gènes NRPS voire des clusters complets de gènes. Par ailleurs, le transfert correct du gène ciblé dans l'hôte hétérologue ne garantit pas le succès de la production du NRP associé. Plusieurs facteurs sont à considérer par la suite, à commencer par le taux de transcription du gène qui peut être amélioré, par exemple en utilisant un promoteur plus approprié. Les étapes de traduction et de repliement de la NRPS doivent également être considérées et peuvent être promues en modifiant les conditions de culture ou par la co-expression d'enzymes chaperonnes natives. De même, afin d'empêcher la dégradation de la synthétase néo-produite, il est important de limiter l'activité protéolytique de l'hôte hétérologue. Une fois la NRPS active obtenue, l'assemblage du NRP peut être limité par une carence en monomère précurseur, ou même une toxicité du composé envers l'hôte choisi. Le premier facteur limitant peut être contourné en clonant les gènes responsables de la biosynthèse des précurseurs, ou en supplémentant ces derniers dans le milieu de culture. La toxicité peut quant à elle être limitée en conférant à l'hôte la capacité de sécréter le composé, par exemple en clonant les gènes identifiés comme étant impliqués dans son transport.

L'approche de production hétérologue a été utilisée notamment afin d'identifier de nouveaux NRPs, d'élucider le mode de fonctionnement de certaines NRPSs, et dans une moindre mesure d'améliorer la productivité. C'est une approche particulièrement pertinente pour produire des composés dont les gènes sont nativement peu ou pas exprimés, chez des organismes pour lesquels les outils génétiques n'ont pas encore été développés. Néanmoins, les rendements en NRPs obtenus actuellement par production hétérologue sont loin d'atteindre les niveaux requis pour une production à plus grande échelle. L'optimisation des conditions de cultures de l'organisme produisant naturellement le NRP, ou l'utilisation de la synthèse chimique, sont les deux approches les plus prometteuses en termes de productivité. Néanmoins, les possibilités offertes par la seconde sont plus limitées dans le cas de composés cycliques, ou contenant un monomère inhabituel ou comportant des modifications difficiles à obtenir par synthèse chimique.

Biotechnology Advances, 2019

Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production

Antoine Vassaux^{a,b}, Loïc Meunier^{a,c}, Micheline Vandenberg^d, Denis Baurain^c,
Patrick Fickers^a, Philippe Jacques^a, Valérie Leclère^b

^aTERRA Teaching and Research Centre, Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d'Agronomie, B5030 Gembloux, Belgium

^bUniv. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, EA 7394-ICV-Institut Charles Viollette, F-59000 Lille, France

^cInBioS-PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liege, Boulevard du Rectorat 27, B-4000 Liège, Belgium

^dTERRA Teaching and Research Centre, Microbiologie et Génomique, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d'Agronomie, B5030 Gembloux, Belgium

Abstract

Fungi are notoriously prolific producers of secondary metabolites including nonribosomal peptides (NRPs). The structural complexity of NRPs grants them interesting activities such as antibiotic, anti-cancer, and anti-inflammatory properties. The discovery of these compounds with attractive activities can be achieved by using two approaches: either by screening samples originating from various environments for their biological activities, or by identifying the related clusters in genomic sequences thanks to bioinformatics tools. This genome mining approach has grown tremendously due to recent advances in genome sequencing, which have provided an incredible amount of genomic data from hundreds of microbial species. Regarding fungal organisms, the genomic data have revealed the presence of an unexpected number of putative NRP-related gene clusters. This highlights fungi as a goldmine for the discovery of putative novel bioactive compounds. Recent development of NRP dedicated bioinformatics tools have increased the capacity to identify these gene clusters and to deduce NRPs structures, speeding-up the screening process for novel metabolites discovery. Unfortunately, the newly identified compound is frequently not or poorly produced by native producers due to a lack of expression of the related genes cluster. A frequently employed strategy to increase production rates consists in transferring the related biosynthetic pathway in heterologous hosts. This review aims to provide a comprehensive overview about the topic of NRPs discovery, from gene cluster identification by genome mining to the heterologous production in fungal hosts. The main computational tools and methods for genome mining are herein presented with an emphasis on the particularities of the fungal systems. The different steps of the reconstitution of NRP biosynthetic pathway in heterologous fungal cell factories will be discussed, as well as the key factors to consider for maximizing productivity. Several examples will be developed to illustrate the potential of heterologous production to both discover uncharacterized novel compounds predicted *in silico* by genome mining, and to enhance the productivity of interesting bio-active natural products.

Keywords: Bioinformatics tools; Fungal cells; Genome mining; Heterologous production; Large gene cluster; Multi-modular enzyme; Nonribosomal peptide; Nonribosomal peptide synthetase; Secondary metabolite

5.1. Introduction

The era of nonribosomal peptides (NRPs) started with the brilliance of Fleming to analyze the unexpected ability of a filamentous *Penicillium* strain to impair bacterial growth. These observations led to the discovery of the penicillin antibiotic, a NRP derived compound which was not identified as such at that time [4]. Indeed, the biosynthesis mechanism of this type of peptides, named thiotemplate mechanism, was elucidated only 42 years later with the description of the Non Ribosomal Peptide Synthetase (NRPS) of the gramicidin antibiotic from *B. subtilis* [15]. Since then, there has been growing interest in the study of this kind of compounds. As a result, NORINE, the first database entirely dedicated to NRPs was established in 2007, starting with 700 NRPs entries [70]. At present, over a thousand of NRPs have been characterized and classified [136]. These NRPs were mostly found in prokaryotes and fungal eukaryotes [19], and are notably not present in metazoa and plantae eukaryotic clades. Few exceptions were reported such as the production of astins in the plant *Aster tataricus*, later refuted and attributed to a fungal endophyte *Cyanodermella asteris* [16]. Likewise, polytheonamide initially isolated from the marine sponge *Theonella swinhoei* [137], was finally shown to be produced by a symbiotic bacteria [138,139].

Nonribosomal peptides, as suggested by their name, are not assembled on the ribosome but by large multi-modular enzymatic complexes that are encoded by gene clusters that often span over tens of kilobases in the genome (**Figure 13**).

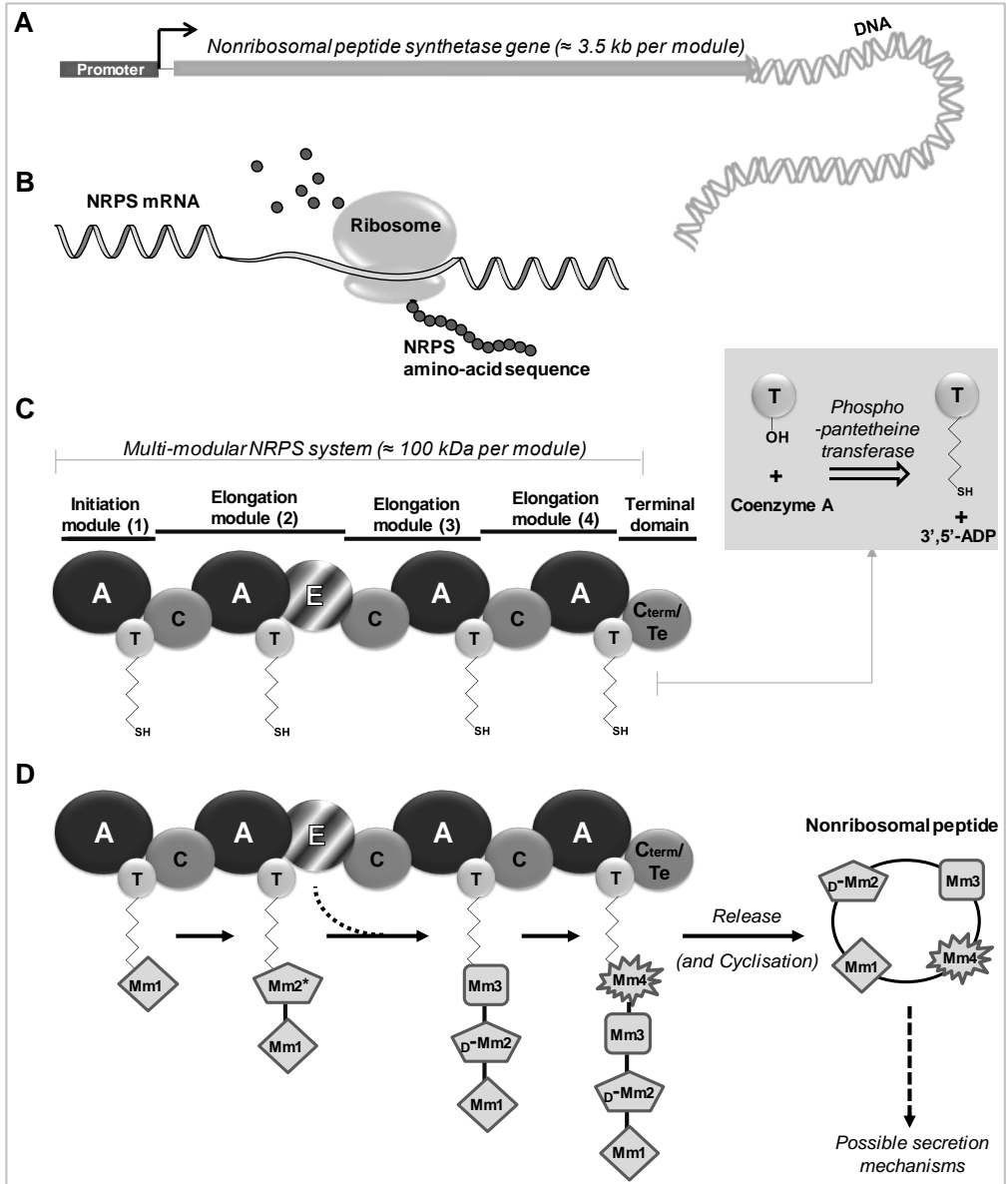


Figure 13: Multiple steps involved in NRP biosynthesis pathway. **A.** Transcription of the NRPS encoding gene. **B.** Translation of the NRPS mRNA. **C.** Folding of the NRPS in multi-enzymatic domains modular organization and addition of phosphopantetheine arm (*holo*-NRPS). Several domains are presented: A: adenylation; T: thiolation; C: condensation; C_{term}/Te: thioesterase or terminal condensation domain (only found in fungal NRPSs); and an example of a potential additional epimerization (E) domain (converting an L-form amino acid to its D-form). **D.** Operating way of NRPS assembly line to generate NRP (Mm: monomer). *Epimerization reaction can occur either before or after condensation reaction.

However, NRP synthetases are synthesized on the ribosome, before being activated by a specific phosphopantetheinyl transferase (PPTase). This PPTase is responsible for the conversion of the *apo*-NRPS into active *holo*-NRPS through the addition of a so-called phosphopantetheine arm onto thiolation domains [41]. NRPSs can be divided into multiple modules, each of them responsible for the incorporation of one peptide monomer. These modules comprise multiple enzymatic domains exhibiting specific functions in the NRP assembly line [70]. The adenylation (A) domain is able to recognize and activate a specific monomer, which is then tethered onto a thiolation (T) / peptidyl carrier protein domain through the phosphopantetheine arm. The flexibility of the multi-enzymatic complex enables the presentation of monomers to the condensation (C) domain which catalyzes the formation of a peptide bond between monomers from two subsequent modules [64]. In NRPSs involved in lipopeptide biosynthesis, a major NRPS class, a starter C domain priming the synthetase enables to acylate the first peptide monomer with the fatty acid part [140]. Some NRPSs also exhibit additional domains responsible for monomer modifications, such as methylation, oxidation and epimerization.

NRPSs predominantly operate according to a linear synthesis, with an initiation module (A-T) able to recognize the first monomer, followed by as many modules (C-A-T) as monomers required to complete the peptide. Nonetheless, some NRPSs rather operate with an iterative or nonlinear organization, and then may not follow the colinearity rule between the number of modules and the size of synthesized peptides [23,141]. The amonabactin synthetase found in *Aeromonas* sp. is a prime example of an atypical assembly line which displays an iterative, alternative and optional mode of synthesis. In this NRPS, the AmoG A-domain activates alternatively a Trp or a Phe, on which are bonded two dipeptides iteratively generated by the AmoE and AmoF domains, and optionally in two out of four cases, an AmoH domain incorporates a glycine between the 2,3-dihydroxy benzoic acid and the lysine of one of the dipeptides [142]. NRPS assembly lines usually end with a termination module comprising a thioesterase domain (TE) responsible for the release, and the cyclization, in most cases, of the final compound [143,144]. In fungi, thioesterase domains can be replaced by condensation terminal domain (C_{term}) ending the NRPS and leading to cyclic NRPS [57]. Alternatively, in NRPSs involved in peptaibols biosynthesis (i.e., NRPS harboring a C-terminal alcoholic group instead of carboxyl group), TE is substituted by a reductase domain releasing the peptide with a C-terminal alcohol [56].

Filamentous fungi are considered as prolific producers of NRPs. A large number of clusters involved in NRP synthesis has been detected for example in *Penicillium thymicola* (17 clusters), *Aspergillus fumigatus* (14 clusters), *Talaromyces islandicus*, *Beauveria bassiana* (13 clusters), and *Pestalotiopsis fici* (12 clusters) [102,145–148]. The high potential of fungal organisms for NRPs production can be exploited for the discovery of novel compounds of interest. The search for novel compounds in fungi is typically achieved following two different approaches: one starting from the biological activities and the other, more recent, from genome sequences. The principle of the former consists in screening samples originating from various environments for

their biological activities. Once the targeted bioactivity is detected, the active natural product is purified and characterized [149,150]. Usually bioactivities are first screened at small-scale and then upscaled to reach higher production rates. However, this strategy is very restrictive as only compounds displaying the screened activity and being produced in sufficient amount to induce an observable activity, will be detected [151]. In contrast, the second approach is based on the use of bioinformatics tools in order to identify the SM gene clusters of interest [151,152]. Frequently, the associated compound is not produced because the related cluster is cryptic and the cultivation conditions were probably not optimal to succeed in its awakening [153,154]. Therefore, the second approach often requires genetic engineering either to activate the expression in the native host (homologous production) or to transfer the biosynthesis pathway in a heterologous host (heterologous production) [155].

In the past few years, microbial genome sequencing and annotation methods have made an incredible leap forward, providing the scientific community a wealth of publicly available fungal genomes [109]. This technological progress has opened new perspectives for the discovery of novel compounds through the bioinformatic approach. The development of novel bioinformatics tools enables, at present, the analysis of multi-modular complexes such as NRPSs, but also polyketide synthases (PKSs) or hybrid complexes PKS-NRPS [156]. The case of polyketides, assembled from acyl- and malonyl- coenzyme A by PKSs, is already well documented and is beyond the scope of this review [157–159]. Nonetheless, this study presents the multiple bioinformatics tools developed to predict the modular organization of NRPSs, and the substrate specificity of their adenylation domains, with an emphasis on the specificities of fungal systems. Bearing in mind that these NRPs identified by genome mining are frequently poorly synthesized under laboratory culture conditions, and taking into account for most of these fungi that adapted molecular tools are lacking, the heterologous expression of their biosynthetic pathways is often a necessary step [160–162]. The heterologous production of NRPs in bacterial organisms such as *Escherichia coli* and *Streptomyces* spp. have been widely reviewed [158,160,163,164], but only few studies have investigated the potential of fungal organisms to produce such foreign compounds. Therefore, this review discusses, through multiple successful examples, the possibility to use fungal cell factories to heterologously express NRPS-encoding genes identified by genome mining. Finally, the multiple key factors to consider obtaining an efficient NRPs heterologous producer are described and then illustrated onto the example of the NRP-derivative penicillin, a multi-decades case of study.

5.2. Computational tools and methods for genome mining

Biosynthetic gene clusters (BGCs) involved in the synthesis of secondary metabolites (SMs) usually contain core synthase genes, regulatory and transport-related elements, and sometimes genes conferring a resistance to the related SM. Most of the developed bioinformatics tools can be applied to the analysis of BGCs related to multiple type of SMs (e.g. NRPS, PKS) coming out from either bacterial or fungal genomic data. Regarding to NRPS analysis, bioinformatics methods have been developed based on the high conservation of the NRPS thiotemplate mechanism and decades of experimental characterization [25,165]. Applications of these algorithms can be used to identify NRPS cluster genes, their subsequent annotation, the prediction of the backbone of the putative NRP produced, and the search for related NRPs. The *in silico* annotation of BGCs relies on the algorithm efficiency, but also on the genome assembly quality, as the quality and the completeness of the annotation depend directly on the input sequences [166]. These two aspects will be further discussed in the following sections.

5.2.1. Recent advances in genome sequencing and annotation

From the mid-2000s, Second Generation Sequencing (SGS) technologies, also named Next Generation Sequencing (including Illumina and Ion Torrent technologies), revolutionized genomics and many other biological fields, by offering low-cost genome sequencing [167]. These developments generated an exponential growth in the genome sequences available in public databases (more than 190,000 prokaryotic genome assemblies available on NCBI databases as of March 2019, <https://www.ncbi.nlm.nih.gov/genome/browse/>) [168]. SGS technologies are based on the principle of a high-throughput method, with a short-read length up to a few hundred bases (e.g., 250 bp for Illumina HiSeq), and usually a low error rate. The drawbacks of short read genome assemblies are related to low contiguity and poor gene completeness [169,170]. As NRPS biosynthetic gene clusters can span several tens of thousands bases - even more when considering “accessory genes” -, BGCs full sequence might be spread into several contigs [171,172]. This causes complex issues for reconstructing BGCs from draft genomes, as well as determining if several unknown gene clusters in an organism are related to only one or several NRPSs.

The recent emergence of Third Generation Sequencing (TGS), with PacBio and Oxford Nanopore technologies, brings new interesting features for BGC annotation. These two methods yield long read length (from a few tens of kb for PacBio, up to 1 Mb for Oxford Nanopore, depending on the quality of the DNA extraction [173]), hence generating high contiguity assemblies [174]. This long read length may present a real advantage in the case of BGC studies, allowing for obtaining complete sequences, and therefore the entire set of involved genes. Yet, in contrast to SGS, TGS displays a lower throughput and a higher error rate [175]. Several software packages, such as Flye [176] and Canu [177], provide efficient methods for TGS data assembly. When the read depth coverage is not sufficient, hybrid strategies taking advantage of the low error rate of SGS can be used for polishing TGS reads or contigs, allowing

for obtaining a high-quality genome assembly [178]. For instance, SPAdes [179] offers hybrid assemblies, while Pilon [180] allows the user to polish TGS assemblies.

Regarding the NRPS identification and annotation, these steps are driven by fast and reasonably accurate bioinformatics methods, which have now become one of the most predominant approaches to support the discovery of novel SMs. This emerging field is promising, as a result of constant effort in software development [156,165,181] and the availability of genomic data. Most methods for annotating NRPSs rely on similarity searches, which is a robust approach as these multi-enzymatic systems are composed of repeated, well conserved domains. The detection of core biosynthetic components is commonly conducted by protein signature searches [181]. These signatures, also called pHMMs (profile Hidden Markov Models) are probabilistic models capturing the versatile information contained in a multiple sequence alignment (i.e., amino acid composition and insertions/deletions per site) to perform specific and sensible annotations [182,183]. Hence, from the prediction of synthetase genes, it is then possible to deduce the NRPS architecture (i.e., domain composition), and search for additional genes surrounding this core (e.g., regulation and transport related genes, tailoring enzymes).

5.2.2. Bioinformatics tools dedicated to NRPS study

Publicly available bioinformatics tools for annotating NRPS started to be devised a decade ago [184], since then, a large number of software packages and methodologies for genome mining analyses has been released (**Figure 14**).

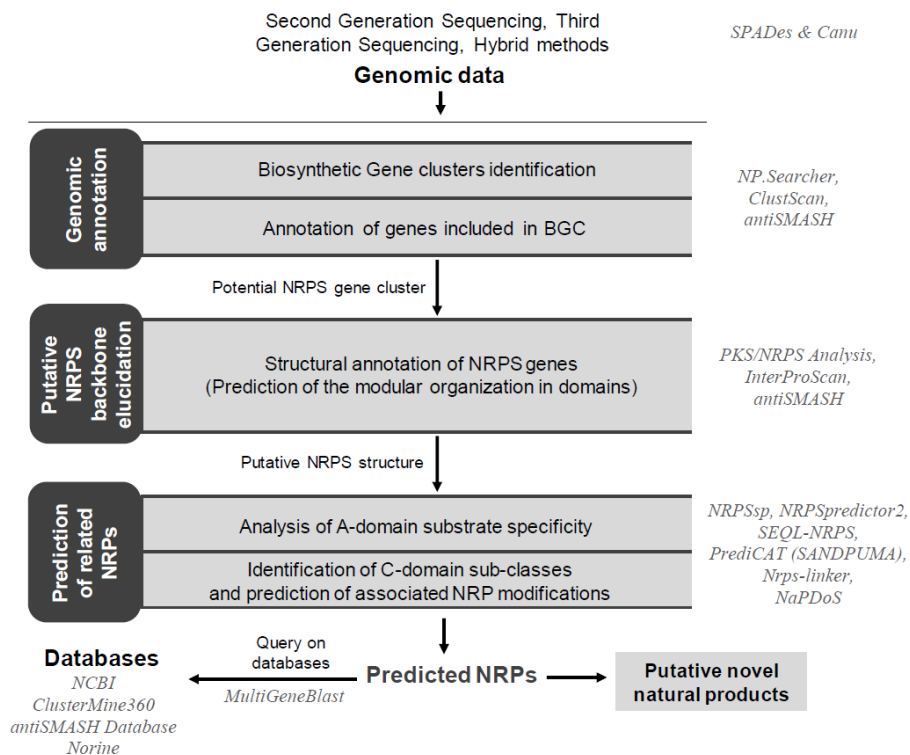


Figure 14: Simplified workflow for novel NRP discovery by genome mining with an overview of the main bioinformatics tools available.

The Secondary Metabolite Bioinformatics Portal represents a very useful online resource (<http://www.secondarymetabolites.org/>) facilitating the exploration of these numerous software packages. It gathers information about the majority of genome mining and annotation tools, as well as databases applied to secondary metabolism investigation [156]. The portion of these tools implying specific functionalities for NRPS and closely related PKS is mostly constituted with generic algorithms applicable for both bacteria and fungi, even if the training of their algorithm was mainly based on bacterial data. For instance, NRPS annotation: NP.searcher [185], ClustScan [186], and antiSMASH [184] are generic tools for the overall annotation of NRPS. The latter is currently the most comprehensive and up-to-date pipeline for BGC analysis, providing a great deal of available functionalities for refining NRPS annotations [156,184]. Among the small portion of tools dedicated to Fungi, SMURF [187] is also a well-known software performing the basic identification of putative synthase and auxiliary enzyme genes. It is also noteworthy that antiSMASH possesses a version for fungal data (i.e., fungiSMASH - <http://fungismash.secondarymetabolites.org>) [188] which adapts the internal gene predictor for an eukaryotic one (GlimmerHMM). In addition, fungiSMASH also offers to apply CASSIS [189], an algorithm aiming the improvement of NRPS/PKS

BGC boundaries delimitation by searching the presence of common regulation pattern in BGC promoters.

In addition to these general NRPS annotation tools, some are entirely devoted to the annotation of NRPS biosynthetic pathways, focusing on specific features of NRPSs and their products. Indeed, these tools aim to predict the peptidyl backbone of the putative NRP produced by analyzing two elements: substrate specificity of adenylation domains and condensation domain subtypes classes.

Substrate selection is based on the specificity of adenylation domains. Determining this specificity is a key element in the structural characterization of the putative synthesized compound. The specificity-conferring code, known as “Stachelhaus or NRPS code”, was the first prediction method based on the identity of ten amino acid residues located in the phenylalanine-binding pocket of PheA (adenylation domain of the first module of the gramicidin synthetase) [33]. These critical residues were demonstrated to be involved in the substrate specificity as they define the binding-pocket recruiting specifically the various monomers. Therefore, with multiple alignments of the already known NRPS A-domains, it was possible to establish a code that correlate these residues to specifically recognized amino acid substrates [190]. Therefrom, several substrate specificity prediction methods were developed such as the pHMM-based approach analyzing the specific protein signatures in the active site [191]. This method was integrated and further developed in NRPSsp (<http://www.nrpsp.com/>) and NRPS-PKS-substrate-predictor (<http://www.cmbi.ru.nl/NRPS-PKS-substrate-predictor/>) [192,193]. Another method, based on a Support Vector Machine approach, is exploited in NRSPredictor2. This software package provides predictions according to the physicochemical properties of the 34 residues in a radius of 8 angstrom around the PheA binding pocket [194,195]. Substrate prediction obtained through the above-mentioned methods are usually accurate when dealing with bacterial data but are still not entirely reliable with fungal data. Three reasons can justify these limitations regarding substrate prediction : (1) low sequence conservation between A-domains (10-40% identity) [196], (2) the existence of different A-domain patterns for a same recognized amino acid [193], or on the contrary, (3) A-domains displaying a relaxed selection of more than one substrate [32,141,197]. Therefore, new strategies have been developed to address the still unresolved challenge of substrate specificity prediction. One recently developed approach, SEQL-NRPS (<http://services.birc.au.dk/seq1-nrps/>), enables to predict substrate specificities based on an alignment-free method detecting occurrences of motifs in domains primary structure [198]. Recently, prediCAT, a phylogenetic-based algorithm enabling to estimate the degree of predictability of A-domain, has been developed and integrated in SANDPUMA [199]. SANDPUMA is a prediction framework combining three methods: (1) a retrained version of the Support Vector Machine models applied in NRSPredictor2, (2) a search for exact matches from active site motifs as in the Stachelhaus *et al.* (1999) approach and (3) a pHMMs analysis like the Khayatt *et al.* (2013) approach. The last antiSMASH version (4.0) modified its previous substrate predictors (Stachelhaus code, Minowa’s pHMMs and NRSPredictor2) to implement SANDPUMA [188].

Condensation domains also exhibit substrate specificity during the elongation reaction, as well as some possible additional activities such as cyclization or epimerization [200]. These domains can be divided in different subtypes (**Table 2**).

Table 2: List of condensation domain subtypes and their associated reactions

Condensation domain subtype	Reaction description	Reference
^L C _L	Bind L-amino acids to L-amino acids	
^D C _L	Bind D-amino acids to L-amino acids	
Starter C	Load a fatty acid as first substrate of the synthesis	[140]
Dual C/E	Combine condensation and epimerization activities	[200]
Epimerization	Invert an L-form amino acid to its D-form enantiomer	
Heterocyclization	Catalyse peptide bond formation, followed by cyclization	
C _{term}	Replace thioesterase activity for the final step of peptide release	[57]
Hybrid C	Bind an amino acid to an aminated polyketide in hybrid PKS-NRPS systems	[200]

The identification of the C domain-subtype is a crucial element for inferring the modifications introduced into the amino acids constituting the final peptide (e.g., identifying D-amino acids or peptide cyclization) [201]. Analysis of condensation domains can be achieved with the NaPDoS web application (<http://napdos.ucsd.edu>), which uses a phylogenetic approach to classify the subtypes of condensation domains and ketosynthase domains from PKSs [200]. Each subtype of condensation domain exhibits specific conserved motifs enabling to efficiently classify them in phylogenetic trees [201]. According to its position into the phylogenetic tree, it is possible by inference to determine the subtype of the analyzed C-domain.

When this additional information regarding core domains specificities has been collected, a quite precise determination of a putative NRP compounds architecture is possible by combining the predicted substrates of adenylation domains, with the modifications related to the predicted condensation classes. Florine, a workflow dedicated to the annotation of NRP compounds, presents a detailed decision diagram to carry out this annotation process [201]. Nonetheless, the mechanism for substrate

selection in NRPS is not fully elucidated and bioinformatics tools are being continuously improved as a result of new discoveries. For instance, the short amino acid regions connecting adjacent modules, called linkers, have been very recently highlighted to also play a role in the determination of substrate specificity. Subsequently to this discovery, a web application just has been released to enable the identification of these inter-modular NRPS linkers (<https://nrps-linker.unc.edu>) [202].

5.2.3. Linking NRPS to related clusters or known products

Once the NRPS architecture is elucidated thanks to the above described bioinformatics tools, it is important to determine if the compound is already described and associated to a characterized NRP (**Figure 14**). A lot of useful tools are available to perform this dereplication steps, but only few of them will be described thereafter. The standard approach for analyzing BGC is to use the Basic Local Alignment Search Tool (BLAST) search program hosted on the NCBI website (<http://www.ncbi.nlm.nih.gov/blast>). This application enables to perform a sequence similarity search relative for instance to sequences listed in the “nucleotide” (BLASTn) or the “non-redundant protein sequences” (BLASTp) collections of the NCBI [203,204]. Alternatively, it is possible to query dedicated databases for BGCs, NRPSs and their products. Regarding BGC pathways, ClusterMine360 is a PKS/NRPS database mainly updated by crowdsourcing, which offers a two-way approach where nearly 1000 gene clusters can be browsed directly from a cluster list or following a search from more than 200 compound families [172]. The antiSMASH database is another relevant resource, encompassing currently a collection of antiSMASH results for almost 25,000 genomes - but mainly of bacterial origin - and offering the possibility to build precise queries [205]. MultiGeneBlast tool enables identifying homologs of multigene modules, such as NRPSs, on the basis of similarity and synteny criteria [206]. MultiGeneBlast is also integrated in antiSMASH, where it suggests some potentially related clusters (or the identical cluster itself) taken from different BGC databases of varying levels of curation. For example, the “KnownClusterBlast” option refers to a curated database containing experimentally characterized clusters, while the “ClusterBlast” option uses an extended database composed mostly of only hypothetical clusters [207].

The only database fully dedicated to NRP compounds, Norine, currently comprises 1190 peptides, including 346 NRPs of fungal origin distributed in 46 families. Norine offers different search modes: (1) an “annotation search” engine, which includes queries with all annotations (e.g., biological activity, molecular weight, monomer count, or producing organism), and (2) a “structure mode” allowing searches based on monomer composition fingerprints, sequence similarity or whole structures. The analysis of predicted peptides permits searching for related analogous compounds already characterized, or to check if some amino acid patterns involved in some biological activities can be spotted [136]. Only a few main databases were mentioned in this section, but a more exhaustive list is provided on the Secondary Metabolite Bioinformatics Portal [156].

5.2.4. Fungal NRPS analysis challenges

Most of the methods and bioinformatic tools available so far were designed both for the analysis of NRPS of fungal and bacterial origins. However, these organisms are taxonomically distant and some differences at the genomic level can be readily identified. Indeed, in comparison with prokaryotic genomes, fungal genomes are characterized by a larger size (i.e., several tens of millions bases), a higher number of chromosomes, a lower gene density, and the presence of introns [208,209]. BGC analysis pipelines, such as antiSMASH, offer an internal gene prediction before applying protein signature searches. Structural annotations (i.e., the prediction of whole gene structure) performed by automated internal gene predictors (e.g., Glimmer3 in antiSMASH [210]) are sufficiently accurate for bacteria, whereas annotation of genes from eukaryotic genomes is usually more complex. Dedicated gene prediction pipelines for eukaryotes often need some external data (gene expression and/or homologous protein data) to support *ab initio* predictions and provide higher quality annotation [166]. As it is essential to have coding sequences correctly annotated (i.e. with intron-exon and intergenic structures well resolved) to accurately predict NRPS gene clusters, fungal genomes are likely to benefit from a preliminary gene prediction, before being processed by BGC analysis tools. The available tools, dedicated to eukaryotic genome analysis, to perform this preliminary gene prediction are depicted in a relevant review [166].

When pioneering predictive models for the substrate specificity of adenylation domains were developed a decade ago [191,194], very few fungal data was available. Therefore, these models provide predictions based mainly on monomers identified as a substrate for the few well-characterized bacterial NRPSs at that time. Consequently, these bioinformatics tools cannot predict with high accuracy the substrate specificity of most of the A-domains of eukaryotic NRPSs. Furthermore, based on the Norine database, it appears that NRPSs of fungal and bacterial origin display only a moderate range of overlapping substrate and incorporate a different set of monomers into the NRPs (**Figure 15**).

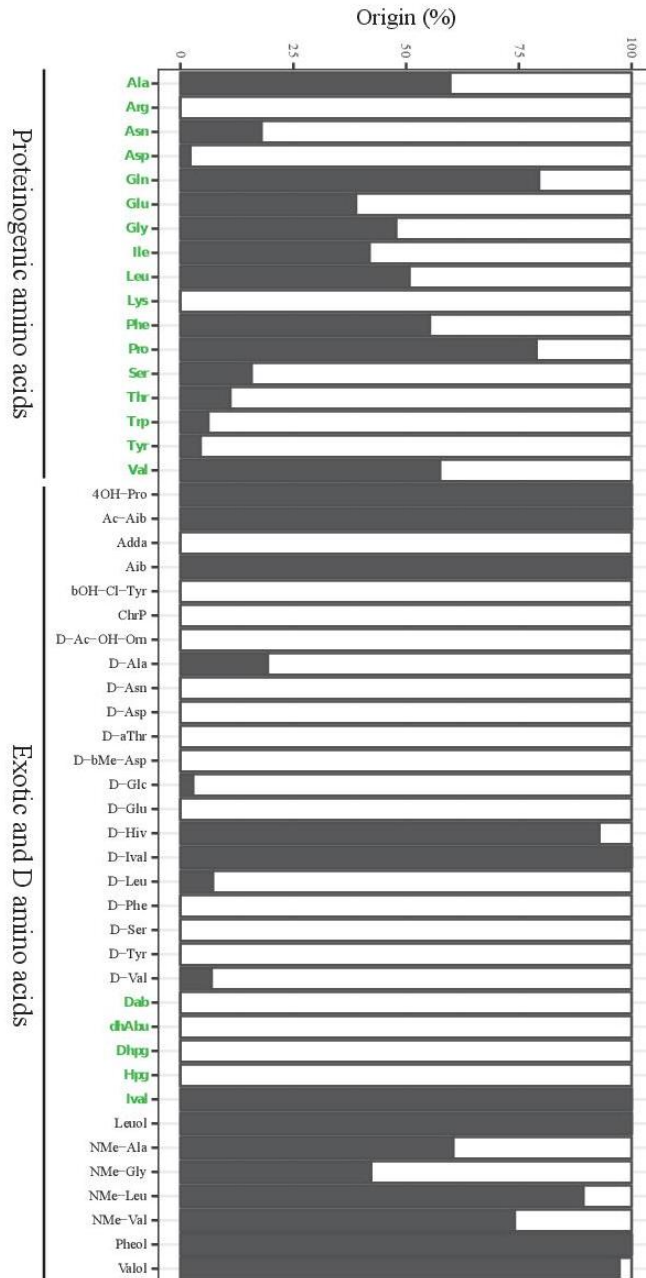


Figure 15: Analysis of the 50 most frequent monomers constituting the nonribosomal peptides classified in the Norine database. Each bar is subdivided into colored parts corresponding to the taxonomic clade from which the resulting NRPs was identified (i.e., white: bacteria, dark grey: fungi). The monomers specifically activated by A-domains which display pHMMs signatures enabling specificity predictions are highlighted in green.

For example, Aib (2-aminoisobutyric acid) and amino alcohols such as Pheol (phenylalaninol) and Leuol (leucinol) seem to be specific to peptaibols produced mainly by fungi. Consequently, some fungal monomers are simply not recognized as no predictive models exist for these substrates. However, predictive models are only applicable if the monomer is directly selected by an adenylation domain and not resulting from the modification of a precursor amino acid. While Aib seems to be directly loaded [211], amino alcohols in NRPs result from a modification of the amino precursors during synthesis of the peptide [56,144]. Determining if the monomer is loaded as such or modified during the synthesis is a complex task as no generalized rules can be applied. For example, D-amino acids are generally the result of an epimerization domain activity on L-amino acids, but, in some cases, D-Ala was reported to be directly loaded [61].

Even if a few software packages, such as NRPSpredictor2 [195], have previously developed models to predict the substrate specificity of adenylation domains, there is still an important need for devising tools enabling to predict more accurately which monomers are incorporated in fungal NRPs. Therefore, future challenges will be to go further in the elucidation of fungal NRPS mechanisms, and to establish novel tools allowing for increased the accuracy of substrate predictions in specific fungal systems, especially regarding to exotic- and D- amino acids.

5.3. Heterologous production of NRPS in fungal host

The number of available fungal genomic sequences in databases is continuously expanding thanks to individual or large-scale collaborative efforts, such as the “1000 fungal genomes project” [212,213]. These available fungal genomes represent a gold mine for novel NRP compounds discovery. Unfortunately, in many cases, the newly sequenced fungi do not (or only poorly) produce these natural products under usual laboratory conditions in synthetic media, which are not always suitable to develop the filamentous fungi or to produce NRPs [162]. One option is to activate or enhance the homologous production in the native host through promoter exchange for instance [214,215]. If this is not possible, due to a lack of molecular tools for example, the heterologous production is one of the best approaches to reach a sufficient yield of production, and enable further characterization of the NRP [11]. In this part, the methodology commonly used for heterologous production of NRP is presented, taking into account different parameters such as the choice of the heterologous host, the genetic tools required for the transformation, the different expression systems used and the methods used for large scale heterologous production of NRP.

5.3.1. Heterologous host choice

The choice of the host is one of the most critical parameters to consider for heterologous gene expression and product synthesis. The selected strain has to present some desirable characteristics, such as accessible and efficient genetic manipulation tools, capabilities to grow at high cell density, ability to secrete the produced compounds, and above all, a codon compatibility with the foreign genes [164]. The well-known bacterial cellular platforms *Escherichia coli*, *Bacillus* sp., and *Streptomyces* sp., which display most of these characteristics, are commonly used for heterologous expression of NRPs. Nonetheless, bacteria exhibit a different codon usage from eukaryotic genes, which must be considered. One option for addressing this difficulty is to provide additional tRNA sequences for missing and rare codons, such as the AUA codon for Ile in *E. coli* [216]. Alternatively, as they are more closely related to filamentous fungi at a taxonomic level, yeasts constitute appropriate hosts to express foreign fungal NRPS genes. Different studies reporting NRP heterologous production in yeasts are presented in (**Table 3**).

Table 3: Previous studies reporting NRP heterologous production in yeasts

Natural product (NRP)	NRPS	NRPS MW (kDa)	Native producer	Heterologous host	Expression system	Ref.
ACV	ACVS (or PcbAB)	450	<i>P. chrysogenum</i>	<i>H. polymorpha</i>	Plasmid pZ4 Plasmid pG4U	[265]
				<i>S. cerevisiae</i>	Plasmid pESC Genomic integration	[311]
					Plasmid pESC	[315]
Fumiquinazoline F	Afu6g12080	450	<i>A. fumigatus</i>	<i>S. cerevisiae</i>	Plasmid pKW1810	[268]
	TqaA		<i>P. aethiopicum</i>		Plasmid pXW55	[57]
Tricyclic ardeemin F-Q	ArdA	432	<i>N. fischeri</i>	<i>S. cerevisiae</i>	Plasmid pXW55	[269]
Beauvericin	BeSYN	352	<i>B. bassiana</i>	<i>S. cerevisiae</i>	Plasmid pXW55	[271] [272]
Bassianolide	BaSYN	348	<i>B. bassiana</i>	<i>S. cerevisiae</i>	Plasmid pXW55	[271] [272]
Brevianamide F	FtmA	319*	<i>A. fumigatus</i>	<i>S. cerevisiae</i>	Plasmid pKW1250	[161]
Acetylaszonalenin	AnaPS	280	<i>N. fischeri</i>	<i>S. cerevisiae</i>	Plasmid pXW55	[57]
Asperlicin C/ D	AspA	276	<i>A. alliaceus</i>	<i>S. cerevisiae</i>	Plasmid pXW55	[270]
Indigoidine	BpsA	141	<i>S. lavendulae</i>	<i>S. cerevisiae</i>	Genomic integration	[219]
D-Phe-L-Leu	TycA	124	<i>B. parabrevis</i>	<i>S. cerevisiae</i>	Plasmid pESC	[277]
	SrfAC	144	<i>B. subtilis</i>			
2,4-dihydroxy 5,6-dimethyl benzaldehyde	ATEG_03630	120	<i>A. terreus</i>	<i>S. cerevisiae</i>	Plasmid pRS414	[275]
Phenguignardic acid	PgnA	104*	<i>A. terreus</i>	<i>S. cerevisiae</i>	Plasmid pESC	[274]
Atromentin	ATEG_03090	101*	<i>A. terreus</i>	<i>S. cerevisiae</i>	Plasmid pESC	[274]
Aspulyinone E	Apva & MelA	92*	<i>A. terreus</i>	<i>S. cerevisiae</i>	Plasmid pESC	[274]

*Estimated on the basis of the amino acid sequence length (110 Da per amino acid)

Among yeasts, *Saccharomyces cerevisiae* displays both a high codon usage compatibility towards foreign fungal genes [217], and a great potential for the secretion of the produced compounds [218]. This host also displays a certain tolerance towards foreign bacterial sequences, as was recently demonstrated with the heterologous production of a NRP initially produced by *Streptomyces lavendulae*

[219]. As it is possible in this yeast to express a NRPS-encoding gene coming from bacteria, which is taxonomically very distant to the host, its potential is even more promising for the heterologous production of fungal NRP. Moreover, the poorly developed secondary metabolism of *S. cerevisiae* is an asset, as it minimizes the competition with homologous SM pathways, and promotes the detection/purification of the targeted compound. Genetic manipulations are also made straightforward by a plethora of molecular tools, and the aptitude of the yeast to efficiently perform transformation-associated recombination (TAR) enabling the cloning of large DNA fragments [153]. However, *S. cerevisiae* lacks a functional PPTase encoding gene, which thus must be introduced into the host [220]. In accordance with its advantages, and despite the risk of obtaining a low rate of production, *S. cerevisiae* is by far the most widespread yeast host for NRP heterologous production. Nevertheless, the non-conventional yeast *Hansenula polymorpha* (also belonging to the Saccharomycetales) can be an alternative choice. This methylotrophic yeast is also easily cultivable, even at large-scale, and strongly regulated promoters are available [221]. Moreover, *H. polymorpha* was successfully used to produce the NRP d-(L- α -aminoadipyl)-L-cysteinyl-D-valine (ACV), a precursor of penicillin and cephalosporin [222]. Another Saccharomycetales, *Pichia pastoris* (now *Komagatella pastoris*) is also an appropriate host to heterologously express large SM clusters. In this host, multiple examples of large multi-modular PKS expression have been reported [223,224]. Nonetheless, to date, no NRP heterologous production has been reported, despite the obvious potential of *P. pastoris* for SM production. Finally, *Schizosaccharomyces pombe* is the only yeast reported to produce naturally a NRP (ferrichrome), which indicates that it expresses a functional PPTase [22,23]. This characteristic would establish this yeast as a suitable host, but the limited genetic toolbox for Schizosaccharomycetales (only distantly related to Saccharomycetales such as *S. cerevisiae*) currently restricts its scope of usage [225].

Despite yeasts, different filamentous fungi taxonomically close to the original host, can be used for the heterologous expression of foreign NRPS genes. Numerous filamentous fungi display interesting intrinsic properties that might make them potentially convenient hosts for heterologous production of NRPs. First of all, they display a highly developed secondary metabolism, with the expression of a large number of NRPs/PKSs, and therefore contain a functional PPTase. In addition, filamentous fungi were numerous times demonstrated to be able to recognize and correctly splice introns from foreign fungal mRNA sequences, bypassing the necessity to purify and clone large introns-free cDNAs from NRP gene cluster [154,226]. Usually, the choice of a taxonomically closely-related host is preferable, since it generally exhibits a good compatibility regarding codon usage, thus facilitating translation [227]. Indeed, in native hosts, mRNA sequences harbor mainly preferred codons, interspersed with scarcer codons in specific zones, so as to slow down the ribosome and enable the pre-folding of the translated proteins [228]. Among the potential fungal hosts, *Aspergillus* spp., such as *A. nidulans*, *A. niger*, and *A. oryzae*, are the more widely used for NRP heterologous production (**Table 4**).

Table 4: Previous studies reporting NRP heterologous production in filamentous fungi

Natural product (NRP)	NRPS	NRPS MW (kDa)	Native producer	Heterologous host	Expression system	Ref.
KK-1	KK-1 NRPS	1,430*	<i>C. clavata</i>	<i>A. oryzae</i>	Genomic integration	[286]
Cycloaspeptide A/E	PscyB	623*	<i>P. sopii</i>	<i>A. oryzae</i>	Plasmid pTYGS-arg	[285]
Ferrirhodin	FSN1	495*	<i>F. sacchari</i>	<i>A. oryzae</i>	Plasmid pTAYAGSarg	[281]
ACV	ACVS (or PcbAB)	450	<i>P. chrysogenum</i>	<i>N. crassa</i>	Cosmid Pcx3.2	[278]
				<i>A. niger</i>	Plasmid pJET1.2	[226]
Bassianolide	BaSYN	350	<i>B. bassiana</i>	<i>A. niger</i>	Genomic integration	[292] [296]
CDP hexa-bassianolide	BaSYN-EnSYN TC ₃	350	<i>B. bassiana</i> <i>F. oxysporum</i>	<i>A. niger</i>	Genomic integration	[296]
Beauvericin	BeSYN	350	<i>B. bassiana</i>	<i>A. niger</i>	Genomic integration	[292] [296]
Octa-beauvericin	BeSYN-BaSYN TC ₃	350	<i>B. bassiana</i>	<i>A. niger</i>	Genomic integration	[296]
[PheLac]-Beauvericin	PSYN M ₁ -BeSYN M ₂	350	<i>R. abscondita</i> <i>B. bassiana</i>	<i>A. niger</i>	Plasmid pVG2.2	[295]
Enniatin	EnSYN	350	<i>F. oxysporum</i>	<i>A. niger</i>	Genomic integration	[141] [296]
Octa-enniatin	EnSYNΔC ₃ -BaSYN C ₃	350	<i>F. oxysporum</i> <i>B. bassiana</i>	<i>A. niger</i>	Genomic integration	[296]
[PheLac]-Enniatin	PSYN M ₁ -EnSYN M ₂	350	<i>R. abscondita</i> <i>F. oxysporum</i>	<i>A. niger</i>	Plasmid pVG2.2	[295]
Brevianamide F	FtmA	319*	<i>A. fumigatus</i>	<i>A. nidulans</i>	Genomic integration	[87]
				<i>A. niger</i>	Plasmid pKW20142	[161]
				<i>N. fischeri</i>	<i>A. nidulans</i>	Genomic integration
Basidioferrin	CsNPS2	275	<i>C. subvermispora</i>	<i>A. niger</i>	Genomic integration	[282]
Asperphenamate	ApmA ApmB	275* 275*	<i>P. brevicompactum</i>	<i>A. nidulans</i>	Genomic integration	[283]
Gliotoxin	GliP	236	<i>A. fumigatus</i>	<i>A. nidulans</i>	Genomic integration	[284]
Microperfurانونe	MicA	103*	<i>A. nidulans</i>	<i>A. niger</i>	Plasmid pSM565	[290]
Aspulvinone E	ApvA MelA	92*	<i>A. terreus</i>	<i>A. nidulans</i>	Genomic integration	[287]
				<i>A. niger</i>	Genomic integration	[288] [289]

*Estimated on the basis of the amino acid sequence length (110 Da per amino acid)

These species are relatively easy to grow on cheap carbon sources in submerged cultures. Moreover, they have been already used for decades at industrial-scale for food and pharmaceutical applications [229,230]. Conversely, *Aspergillus* spp. exhibit an extensively developed secondary metabolism, which may constitute a drawback. Indeed, the subsequent detection and purification of heterologous compounds are complicated due to the presence of a multitude of natural products. Among those species, *A. oryzae* displays the most limited endogenous secondary metabolism, providing a trade-off to the drawback just described [231]. Moreover, this species is GRAS (generally recognized as safe), especially as concerns the absence of production of aflatoxin, a common mycotoxin found in most of *Aspergillus* species [232]. Lastly, numerous genetic tools, such as selectable markers and various strong promoters, are available to easily manipulate *A. oryzae* [233]. This is partly because of its close relatedness with the genetic model *A. nidulans*, thus resulting in the compatibility of genetic tools [11].

5.3.2. Genetic toolbox for the cloning of genes involved in NRP biosynthesis

Once an appropriate host has been selected, the genes involved in the biosynthesis of the NRP have to be transferred from the natural producer to the heterologous host. In the case of NRPs heterologous production, this step can be particularly challenging for two reasons : (1) NRPS-encoding genes can encompass tens of kilobases which are not easily amplifiable by classical PCR operation, (2) NRPs synthesis often relies not only on the expression of the NRPS encoding-gene but requires the transfer of a set of genes which are generally grouped within the same cluster in the genome of the natural producer. This section will depict some of the solutions to overcome these issues.

5.3.2.1. Cloning of NRPS long sequences

As cloning NRPS genes encompassing up to tens of kilobases can be challenging, several methods have been developed to enable the reconstitution of a long fragment from multiple amplified PCR fragments (e.g., overlap extension PCR, Gibson assembly, USER cloning) (**Figure 16**).

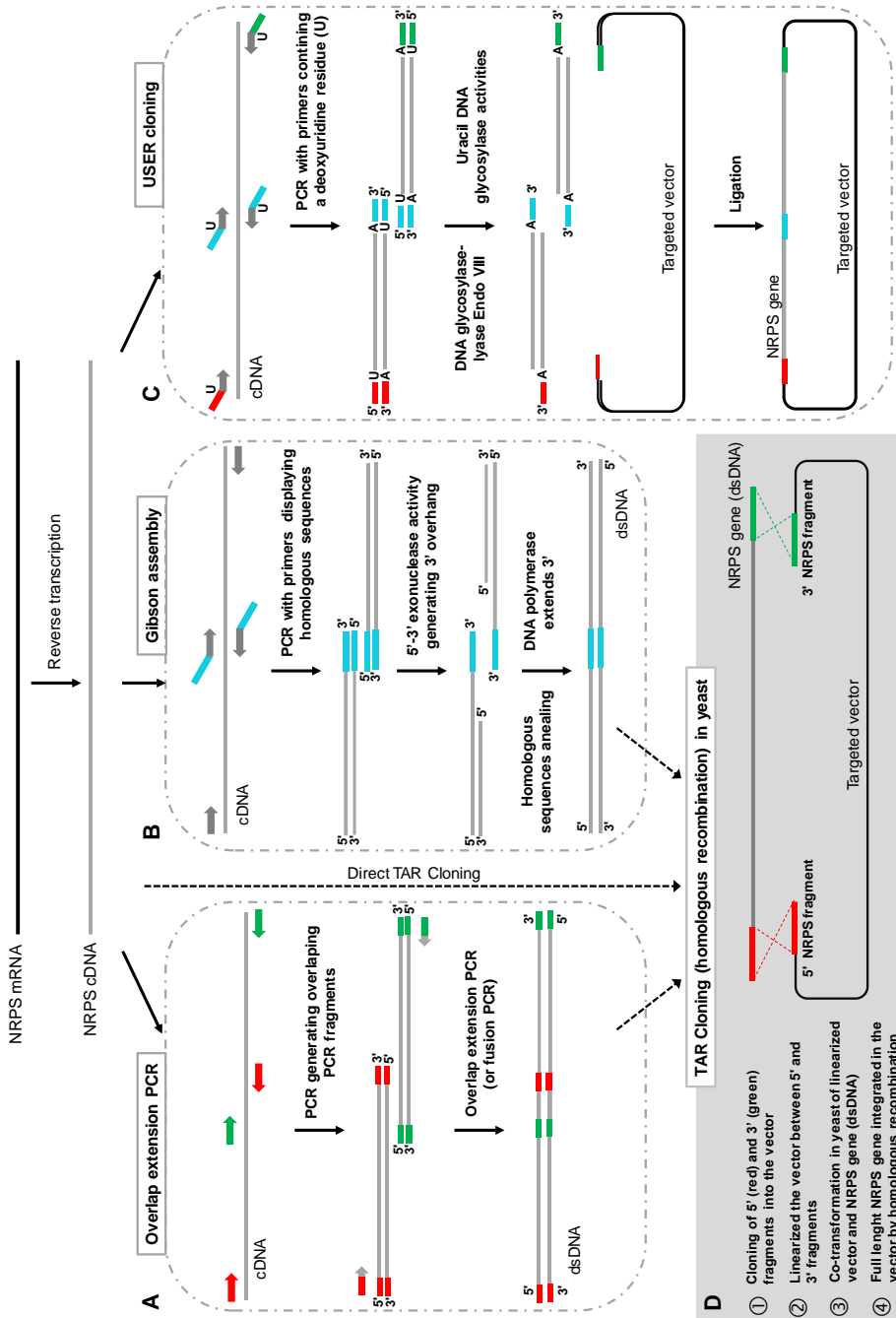


Figure 16: Description of the four main cloning methods enabling the reconstitution, from multiple amplified PCR fragments, of long-length NRPS sequences into a targeted vector: **A.** Overlap extension PCR approach, **B.** Gibson assembly, **C.** USER Cloning, and **D.** TAR Cloning (homologous recombination).

It is advisable to first perform a reverse transcription to generate a single-strand NRPS cDNA, which comprises the intron-free open reading frame, from a pool of total mRNA. The overlap extension PCR approach (**Figure 16A**) permits then the assembly of multiple overlapping fragments, obtained via standard PCR, into a full length double-strand NRPS DNA [234,235]. The Gibson assembly method (**Figure 16B**) is another approach enabling the fusion of multiple DNA fragments. To set-up this method, the multiple PCR fragments are first amplified by standard PCR with primers displaying homologous sequences (approximately 40 bp). The isothermal one-step assembly of the fragments exhibiting terminal overlapping sequences is possible through the joint action of a proofreading DNA polymerase, a Taq DNA ligase and a 5' T5 exonuclease. First, the 5' T5 exonuclease generates single-stranded 3' DNA overhangs by withdrawing nucleotides from the 5' ends. The generated homologous overhangs can then anneal, and the full-length double-stranded DNA be synthesized through the action of the DNA polymerase and the Taq DNA ligase. This system can generate very large DNA fragments, up to several hundreds of kilobases, with high efficiency [236]. Likewise, the Uracil-Specific Excision Reagent (USER) cloning (**Figure 16C**) is also an efficient approach to assemble multiple DNA fragments. To implement USER, a deoxyuridine residue (U) is first included through the designed primers in each PCR fragment. Subsequently, the U residue is cleaved by a DNA glycosylase-lyase Endo VIII and a uracil DNA glycosylase, thus generating a 3' overhangs, enabling a directional assembly of PCR fragments in a linearized target plasmid [237,238].

Once the long fragment is isolated (e.g. gDNA, cDNA) or reconstituted through one of the above-mentioned approaches, TAR cloning is the most commonly used method for its transfer into cosmids or plasmids. Yeasts, especially *S. cerevisiae*, have the ability to perform homologous recombination resulting in the incorporation of the long fragment into the vector. The TAR cloning approach (**Figure 16D**) implies first the cloning of two fragments homologous to the 5' and 3' parts of the long-length sequence, separated by a restriction site, into the vector. The vector comprising 5'-3' parts is then linearized on the above-mentioned restriction site and co-transformed in yeast with the full-length fragment [239]. The denomination “direct TAR Cloning” is used when gDNA or cDNA, eventually prior enzymatically digested, are directly used as a full-length fragment in the co-transformation [240]. On the contrary, if the full-length fragment was assembled by overlap extension PCR before to be inserted in the vector through homologous recombination, the method is named ExRec (overlap extension PCR – yeast homologous recombination) [217,241].

However, the insertion of long NRPS sequences into plasmids leads to a raise in the overall plasmid size, which is dramatically affecting the subsequent cloning efficiency. The maximum insert size can vary from one plasmid to another, but it is generally accepted that the insert size can reach up to 25 kb. Beyond this size, the use of cosmids, which are plasmids containing a phage sequence, enables to clone larger sequences (30-50 kb) in a more stable way [242,243]. In filamentous fungi, an alternative method for cloning large SM clusters via Fungal Artificial Chromosomes (FACs) has been developed [244,245]. An equivalent artificial chromosome system

is also available in yeasts (YACs) [246]. The use of artificial chromosomes has enabled the direct cloning of DNA fragments up to 300 kb [240].

Alternatively, a very efficient way to isolate a pre-identified NRPS gene is to take advantage of *de novo* synthesis technology to re-assemble the constituting oligos. This method was already applied to synthesize the bottromycin NRPS gene with the purpose of expressing it in *Streptomyces* species [247], and could be similarly implemented in a fungal host. Nonetheless, even if the cost of this technology is continuously decreasing, it remains the most expensive way to obtain a long DNA sequence [248].

5.3.2.2. Cloning of a large number of genes

Beside NRPS genes, numerous genes such as those encoding a PPTase, tailoring enzymes, or transporter-associated genes, are required to successfully produce a bioactive NRP or NRP derivative. The cloning of many genes can imply the synthesis of several plasmids which is laborious, costly, and problematic in case only few selection markers are available for the chosen host [249,250]. Several solutions particularly well-suited to NRP heterologous production have been proposed to alleviate this issue. One possible strategy consists of assembling the genes in the same plasmid to create a multiple-gene expression plasmid [217]. This approach was successively applied to heterologously express the aspyridone PKS-NRPS gene and its tailoring enzymes in *A. oryzae*. The aspyridone cluster identified in *A. nidulans* comprises the *adpA* PKS-NRPS gene and six genes encoding tailoring enzymes (*adpB*, *adpC*, *adpD*, *adpE*, *adpF*, and *adpG*). A plasmid was synthesized to enable the simultaneous expression of combinations of four genes under the control of four promoters (P_{amyB} , P_{adh} , P_{gpdA} and P_{eno}) in a single plasmid [251]. The golden gate cloning method can facilitate to a considerable extent the process of cloning multiple transcription units (promoter – gene – terminator) in a single expression vector. This strategy relies on the use of type II restriction enzymes that can generate a specific 5' or 3' DNA overhangs of 4 nucleotides depending of the enzyme considered. Subsequently to the enzymatic digestion, two fragments with appropriate cleavage site can be ligated in a one-step restriction-ligation [252]. Based on the golden gate method, easily tunable vectors have been designed with proper cleavage sites enabling to introduce and interchange DNA elements (e.g. promoters, coding sequences, terminators, protein tag sequences, resistance genes) as mere building blocks. These kinds of DNA assembling platforms have already been designed for several potential hosts such as the filamentous fungus *Ustilago maydis* [253], and the yeasts *S. cerevisiae* [254,255], *S. pombe* [256], and *Y. lipolytica* [257]. Alternatively, viral 2A peptide sequences can also be used to express several genes under the control of the same promoter in a single plasmid. These sequences, interspaced between each ORF, are cleaved during protein translation, hence releasing the proteins from each other [226,258].

However, if the use of multiple plasmids and genomic integrations is required, methodologies allowing for recycling of selection markers have been developed in some common fungal hosts. The loss of *URA3* selection gene in *S. cerevisiae*, can be

achieved by cultivation on 5-fluoroorotic acid plates [259,260]. The cre-lox system also enables selection marker recycling in yeast such as *S. cerevisiae* and *P. pastoris* [261,262]. A recycling system is also operational in *A. nidulans*, where it is possible to replace the marker of a previous transformation by cloning into the first selection gene, a second construction with a different marker [263,264].

5.3.3. Fungal expression system

5.3.3.1. Yeast heterologous expression systems

The first study reporting a successful heterologous expression of a fungal NRPS in yeast was related to the production of ACV in *H. polymorpha* [265]. This example will be discussed in detail in section 4. In contrast, for all subsequently reported heterologous expressions of NRPS in yeast, *S. cerevisiae* was selected as a host (**Table 3**), for example with the fumitremorgin biosynthetic pathway from *A. fumigatus* [161]. The precursor of fumitremorgin is the NRP brevianamide F, resulting from a condensation of L-tryptophan and L-proline by FtmA NRPS. This compound is then prenylated by FtmB to form tryprostatin B, before being further modified by FtmC, FtmD, FtmE, and FtmF leading to spirotryprostatin A or B. The adopted approach for the cloning of the 8-kb *ftmA* gene was to take advantage of the capability of *S. cerevisiae* to perform TAR cloning [239]. With the purpose of removing introns, sometimes not or mis-spliced by heterologous yeasts, the total mRNA from the native producer was extracted and converted into cDNAs by a reverse transcriptase. After homologous recombination, the intron-free *ftmA* was inserted into a plasmid designed for expression in *S. cerevisiae* under the control of the strong galactose-inducible GAL1 promoter. After the cloning of *ftmB*, the NRP derivative tryprostatin B was detected at 100-fold higher concentration than in the native producer *A. fumigatus* (35.6 and 0.383 mg.l⁻¹, respectively) [161]. This significant increase in tryprostatin B titers is a prime example for illustrating the potential of the heterologous expression approach to improve NRP productivity.

A heterologous expression approach has also been used to experimentally assign a NRPS cluster identified on the basis of genome mining to the synthesis of a given NRP. In *A. fumigatus*, 14 gene clusters were identified to be involved in the biosynthesis of NRP compounds [266]. Among them, Afu6g12080 was identified and predicted to be involved in fumiquinazoline biosynthesis [267]. Using a similar approach to that taken with *ftmA*, total mRNA was extracted from *A. fumigatus* and converted into cDNAs by a reverse transcriptase. From this cDNA, multiple overlapping PCR fragments with homologous ends (<3 kb) of the introns-free Afu6g12080 NRPS gene were amplified and assembled together using overlap extension PCR methodology. The resulting 11.8-kb fragment was cloned under the control of the strong constitutive P_{tef1} promoter in pKW1810, a *S. cerevisiae* expression plasmid, by TAR cloning. The coding sequences for Flag and 6xHis were added, respectively, to the 5' and 3' ends of the NRPS gene. The presence of tags attached to protein ends facilitates the further detection and purification of the produced NRPS. In this study, the authors were able to confirm the expression of

Afu6g12080 by Western blot using antibodies directed against 6xHis and Flag tags. Subsequently to the addition of the *npgA* gene encoding a PPTase, fumiquinazoline F (Anthranilic acid [Ant] - L-Trp - L-Ala) was purified from the culture broth of the constructed strain (0.4 mg.l^{-1}). Moreover, an analog compound containing a hydroxylate-Ant was also identified, highlighting the plasticity of the first A-domain of the NRPS encoded by Afu6g12080 [268].

The Afu6g12080 NRPS from *A. fumigatus* is very similar to TqaA, another trimodular NRPS identified in *Penicillium aethiopicum*. With the purpose of devising a ready-to-use host for NRPS/PKS heterologous production, Gao and coworkers have introduced the *npgA* PPTase encoding gene into the genome of *S. cerevisiae* (Strain BJ5464-NpgA). Meanwhile the *tqaA* cDNA, devoid of introns, was obtained by reverse transcription on mRNA extracted from *P. aethiopicum*. Subsequently, TAR cloning enabled the integration of the 5'-Flag modified NRPS gene in the yeast expression plasmid pXW55 under the control of the strong constitutive ADH2 promoter. The heterologous expression of *tqaA* was demonstrated through successful purification of the 450-kDa synthetase with anti-Flag antibody affinity chromatography followed by gel filtration. Afterward, the activity of the synthetase was confirmed with the detection of fumiquinazoline [57].

The strategy developed to express *tqaA* [57] was repeated several times to heterologously express NRPS genes in the BJ5464-NpgA strain with the same plasmid. The *ardA* NRPS gene from *Neosartorya fischeri*, was in such manner expressed in this strain, leading to tricyclic ardeemin FQ production [269]. From the same filamentous fungi, *anaPS*, involved in the biosynthesis of a precursor of acetylazonalenin, was heterologously expressed to investigate the role of the terminal condensation domain in NRP cyclization [57]. The expression of the 276-kDa bimodular synthetase AspA from *Aspergillus alliaceus* in *S. cerevisiae* revealed the iterative mode of synthesis of the first module as it led to the synthesis of a tri-peptide Ant-Ant-Trp [270]. Likewise, the production of fungal anticancer cyclooligomer depsipeptides, initially produced by *B. bassiana*, was reconstituted in *S. cerevisiae* through cloning of the gene encoding the bassianolide synthetase (348 kDa) and the beauvericin synthetase (352 kDa) [271,272].

Despite the opportunities offered by TAR cloning methodology, the insertion and correct expression of large sequences, such as NRPS genes, remain difficult in numerous cases. One potential approach to address this challenge is to dissect and clone the synthetase gene into separate modules. The feasibility of module dissection was first been highlighted with the aspyridone synthetase AdpA, encompassing a single NRPS module fused to the C-terminal part of a PKS. The sequence encoding this module was isolated from *A. nidulans* and expressed in *S. cerevisiae*. Subsequently the heterologous AdpA module was purified, prior to confirming the correct enzymatic activity by *in vitro* production of acyltetramic acid in co-incubation with the PKS part [273]. Conversely, several mono-module NRPS-encoding genes from *Aspergillus terreus* were expressed in the multi-copy pESC-URA plasmid: *apvA-melA*, *pgnA*, and ATEG_03090, respectively involved in the biosynthesis of aspolyinone E, phenguignardic acid, and atromentin [274]. The cryptic mono-module

NRPS encoding gene ATEG_03630 from the same filamentous fungi was similarly awakened by heterologous production, leading to the production of 2,4-dihydroxy-5,6-dimethyl benzaldehyde [275].

NRPSs encompassing more than one module can be dissected and each module cloned separately. Nonetheless, some key factors must be considered. Primarily, the position where the NRPS can be dissected is crucial. In the case of the bimodular AspA, the A₁-T₁-C₂ tridomain can be solubly expressed unlike the bidomain A₁-T₁. Similarly, the second module was soluble in the four domains C₂-A₂-T₂-C_t form and not in the A₂-T₂-C_t form [270]. Recently, Bozhüyük *et al.* (2019) have suggested an alternative fusion site area between the two sub-domains of the C-domain, where NRPSs can be dissected. Authors postulated that the self-contained catalytically active unit can be redefined as C_{Asub}-A-T-C_{Dsub}, where C_{Asub} is the sub-domain accepting the growing peptide chain in the N-terminal C-domain and C_{Dsub} is the sub-domain providing the peptide chain in the C-terminal C-domain. Genetic engineering enables, by exchanging these functional units between different NRPSs with similar domain composition, to generate chimeric NRPSs assembling novel NRPs. Up to now, these units have been employed only to reconstitute entire chimeric NRPSs, and there is no evidence that they can be expressed separately in a soluble form [276]. A second point to consider for the dissection of NRPSs is related to the presence of intermodular linkers, which are short amino acid regions connecting covalently subsequent modules. The importance of these short sequences has recently been highlighted between modules one and two of the beauvericin and bassianolide synthetases [272]. The dissection of these synthetases by cloning module 1 and module 2+3 into two distinct plasmids without disrupting the pre-identified linker sequence led to an effective production of beauvericin and bassianolide. In contrast, the same experiment without the linker did not lead to any production of cyclooligomer depsipeptides. In addition to linkers, there are also short amino acid sequences, called COM domains, enabling noncovalent association of the different proteins when the NRPS is composed of several proteins such as in the tyrocidine synthetase or the surfactin synthase. These COM domains are essential when different fractions of the NRPS are expressed separately but can also be exploited to generate chimeric NRPS systems. This was demonstrated with the co-expression of the first module of the tyrocidine NRPS (TycA) from *Brevibacillus parabrevis* with the third module of the surfactin synthase (SrfAC) from *Bacillus subtilis* [277]. TycA was expressed with its native COM^D domain, whereas SrfAC was fused with the COM^A domain, compatible with COM^D, isolated from TycB module. Co-expression in *S. cerevisiae* of these two complexes into distinct pESC plasmids lead to the expected production of ^DPhe-L^LLeu.

5.3.3.2. Filamentous fungi heterologous expression systems

With the exception of the *pcbAB* NRPS gene (involved in ACV synthesis), expressed in *Neurospora crassa* [278], *Aspergillus* spp. were predominantly used to heterologously express NRP compounds (**Table 4**). One of the earliest reported NRP heterologous expressions in filamentous fungi was related to brevianamide F from *A. fumigatus*. The brevianamide NRPS gene, also later expressed in *S. cerevisiae* as described above, had previously been identified through heterologous expression in *A. nidulans* [87]. In order to prevent mis-amplification, the 8.7-kb *ftmA* gene encoding the bimodular NRPS was amplified from *A. fumigatus* as two distinct fragments (5' and 3'). The fragments were assembled together, cloned under the control of the alcohol dehydrogenase promoter (P_{alcA}), and integrated in the genome of the host strain using *pyrG* as a selection marker. In *A. nidulans*, P_{alcA} is a very strong promoter induced by ethanol or L-threonine, and repressed in the presence of glucose [279]. Cultivation under induction conditions led to brevianamide F synthesis with a titer of 45 mg.l⁻¹, which represents a 3,500 fold increase as compared to native strain *A. fumigatus* Af2936. Later, *ftmA* gene, assembled through ExRec overlap-extension PCR-yeast homologous recombination (**Figure 16**), was also expressed in *A. niger* [161]. The FtmA synthetase from *N. fischeri*, displaying 88% identity at the amino acids level with that from *A. fumigatus*, was also successfully expressed in *A. nidulans* [280]. *FtmA* gene under the control of the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (P_{gpdA}) was inserted into the genome of *A. nidulans*, resulting in a production of brevianamide F ranging between 9.7 and 36.9 mg.l⁻¹, depending on the mutant considered. Interestingly, the productivity from one positive mutant to another displayed large variations (from 1 to 4 magnitudes) [87,280]. This can be due to the ectopic integration of the gene in variable regions of the genome.

Numerous fungal heterologous expression studies were conducted to decipher the mode of synthesis of NRPSs. For example, the typical mode-of-operation of two siderophore synthetases has been highlighted by heterologous expression in filamentous fungi. For this purpose, *FSN1* from *Fusarium sacchari* and *CsNPS2* from *Ceriporiopsis subvermispora* were respectively expressed in *A. oryzae* and *A. niger* [281,282]. The *FSN1* gene is encoding a NRPS involved in the synthesis of ferrirhodin, whereas *CsNPS2* is related to the biosynthesis of a type VI siderophore. In both cases, the NRPSs were operating through an iterative mechanism with three additional thiolation-condensation domains receiving the same monomer (N₅-acetyl-N₅-hydroxy-L-ornithine) from a single upstream located adenylation domain. Likewise, through heterologous expression in *A. nidulans*, the synthesis of asperphenamate has been demonstrated to be dependent on the joint action of two distinct NRPSs: ApmA and ApmB. Indeed, the NRP produced by ApmA, N-benzoylphenylalaninol, binds onto the monomer attached to ApmB to generate subsequently the asperphenamate compound [283]. On the other hand, there are some studies reporting identification and confirmation of NRP biosynthetic pathways through heterologous expression. In the case of gliotoxin, a toxin synthesized by *A. fumigatus*, the related gene cluster was known but the role of each gene had not been elucidated [284]. NRPS gene *gliP* expressed under the control of P_{alcA} in *A. nidulans*

was identified to be involved in the first step of gliotoxin biosynthesis with the assembly of NRP-type precursors: cyclo-(L-phenylalanyl-L-seryl) and cyclo-(L-tryptophanyl-L-seryl). Likewise, by heterologous expression in *A. oryzae*, *pscyA* and *pscyB* genes from *Penicillium sopii* have been demonstrated to be the only two genes required for the biosynthesis of cycloaspeptide. Methylated amino acids constituting this NRP are provided through the activity of an N-methyltransferase encoded by *pscyA*, while the peptide assembly is performed by a 5-modules NRPS encoded by *pscyB* [285]. Another reported example was the case of KK1, a NRP natively produced by *Curvularia clavata*, displaying anti-fungal activity against the plant pathogen *Botrytis cinerea*. The 40-kb gene cluster potentially involved in KK1 synthesis was split into two fragments and cloned into *A. oryzae*. The transformed strain was able to produce KK1, demonstrating the role of this gene cluster in the NRP synthesis [286].

Some mono-modular NRPSs have also been heterologously expressed in filamentous fungi. The *melA* gene from *A. terreus* was cloned into *A. nidulans* [287] and *A. niger* [288,289] resulting in the production of aspulvinone E. Similarly, following the cloning of *micA* obtained from *A. nidulans*, the transformed *A. niger* strain was able to produce microperfuranone [290].

A. niger was also used to heterologously express several cyclodepsipeptides, high value NRP compounds, with anticancer, antimicrobial and anthelmintic properties [291]. Enniatin, beauvericin, and bassianolide belonging to this group of compounds are assembled by iterative bimodular synthetases able to link D- α -hydroxy acid with L- amino acids. The *esyn1* NRPS gene from *Fusarium oxysporum* was expressed in *A. niger*, leading to the production of enniatin at initial un-optimized titer of 1 mg.l⁻¹ [141]. Beauvericin and bassianolide were also heterologously produced after the cloning of the corresponding *bbBeas* and *bbBsIs* synthetase genes from *B. bassiana* [292]. In these two studies, a promising *A. niger* heterologous expression host was developed based on the Tet-on bacterial-fungal hybrid expression system. Tet-on system encompassed an rtTA2S-M2-dependent promoter controlling the expression of the NRPS gene and a tetracycline-dependent transactivator rtTA2S-M2. Addition of a tetracycline derivative, doxycycline (Dox), in the medium activated the system and induced the expression of the gene. The expression strength is directly correlated to the Dox concentration added, thus making Tet-on tuneable and very efficient, with expression levels equivalent to those obtained with the widely used strong constitutive promoter *P_{gpdA}* [293]. In addition, the impact of the number of *bbBeas* gene copies integrated in *A. niger* genome on the beauvericin productivity was also investigated [292]. The strain harboring a tandem gene copy produced approximately two-fold higher titers of beauvericin than the strain harboring a single copy, under similar culture conditions (628 versus 294 mg.l⁻¹).

The typical iterative mechanism and the high homology of bimodular cyclodepsipeptide synthetases were also exploited to assemble hybrid NRPSs. In the case of closely related synthetases, inter-module swapping is one of the most effective approaches for altering the incorporated monomer, and thus generating a novel putatively bioactive compound [294]. For this purpose, the first module of a

synthetase is combined with the second module of another system in order to generate new non-natural nonribosomal peptides. Studies reporting this inter-module swapping between cyclodepsipeptide synthetases have employed the above described Tet-on expression system in *A. niger* [295,296]. Module 1 of PSYN NRPS (from *Rosellinia abscondita*), activating either D-Lac or D-Phe, was fused to module 2 of EnSYN (from *F. oxysporum*) and BeSYN (from *B. bassiana*) activating L-Val and L-Phe, respectively. The subsequently transformed *A. niger* colonies were able to produce six novel compounds, but with lower yields than for native cyclodepsipeptides. The lower yields observed can also be related to a bottleneck in substrates for different derivatives [295]. Using a similar approach, after screening hybrid NRPS resulting from several combinations of modules from different cyclodepsipeptide NRPSs (i.e., EnSYN, BeSYN, and BaSYN) in *E. coli*, the most promising combinations were expressed in *A. niger* [296]. The combinations EnSYN-BaSYN.T.C₃, BeSYN-BaSYN.T.C₃, BaSYN-EnSYN.T.C₃, and BaSYN-BeSYN.T.C₃ were successfully expressed, leading to high titers of hybrid cyclodepsipeptides: octa-eniatiin B (4 mg.l⁻¹), octa-beauvericin (10.8 mg.l⁻¹), and hexa-bassianolide (1.3 g.l⁻¹). Module and domain swapping between different PKS/NRPS is an innovative approach to generate new compounds, difficult to obtain by chemical synthesis, and displaying interesting biological activity [297]. With this approach, 6 novel compounds were also produced in *Fusarium heterosporum*, after 57 swapping combinations with 34 modules from various fungal PKS-NRPS [298]. Other notable examples of chimeric compounds include: niduchimaeralin A/B in *A. nidulans* by swapping the PKS-NRPS CcsA from *Aspergillus clavatus* with the Syn2 from *Magnaporthe oryzae* [299], and predesmethylbassianin A / pretenellin A in *A. oryzae* by swapping between the PKS-NRPS DMBS and TENS from *B. bassiana* [300].

Likewise, hybrid PKS-NRPS genes have also been heterologously expressed in filamentous fungi to produce compounds of interest. For example, tennelin was produced in *A. oryzae* after expression of *tenS* PKS-NRPS from *B. bassiana* under the control of the starch-inducible *amyB* promoter (P_{amyB}), leading to up to 5-fold higher yields than in the native fungus [301,302]. Under the control of the same promoter, a PKS-NRPS from *Chaetomium globosum* was expressed in *A. nidulans*, enabling the production of the potential anti-HIV Sch 210972 [303]. Once again, P_{amyB} was employed to express the 12.4-kb *ACE1* PKS-NRPS from *M. oryzae* in *A. oryzae*, leading to magnaportheapyrone production [304]. Similarly, the aspyridone biosynthetic pathway (*adpA* gene cluster) was reconstituted in *A. oryzae* by cloning the genes of *AdpA* cluster from *A. nidulans* [251]. Additionally, two successive PKS and NRPS genes (*glpks3-glnrps7*) in one cluster of *Glarea lozoyensis* were also successfully expressed as a polycistronic mRNA in *A. nidulans* to produce xelolozoyenone [305]. The biosynthetic pathways of astechrome and macrolactone valactamide were reconstituted also in *A. nidulans* through fungal artificial chromosomes.

5.3.4. Key factors to scale up heterologous production of NRPs

Multiple key factors have to be considered to succeed in NRP heterologous production and to increase the productivity. Each step, from the cloning and the

transcription of the long heterologous gene to the functional synthetase, can be a barrier to the production (**Figure 17**). Some of these limiting steps, and the possible solutions to overcome them, will be discussed in this section.

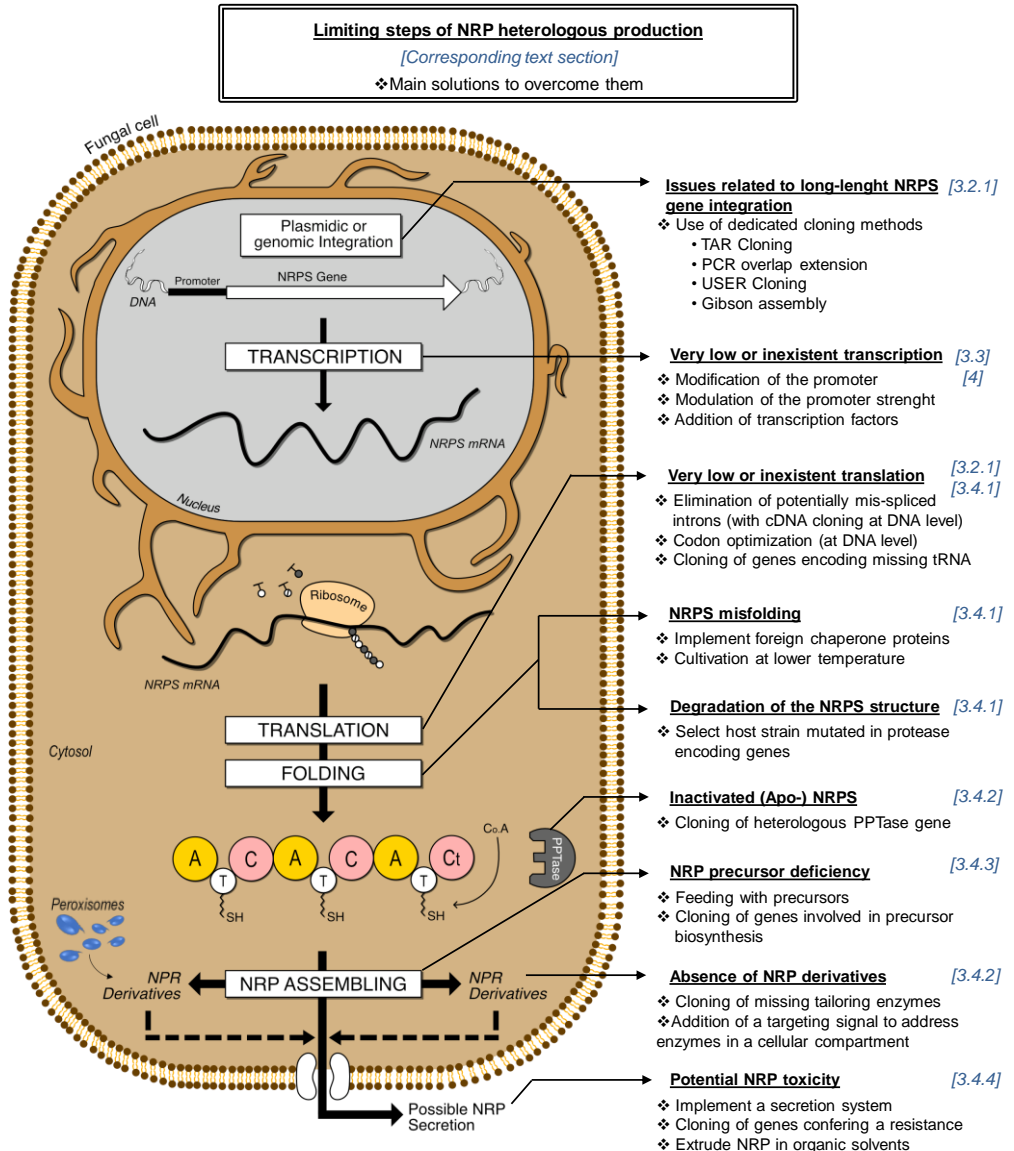


Figure 17: Overview of the limiting steps of NRP heterologous production and of the main solutions to overcome them and consider upgrading production rates.

5.3.4.1. Incompatibility of the host with foreign NRPS nucleic and amino-acid sequences

The incompatibility of the foreign nucleic acid sequences with the host machinery is a major cause of failure of heterologous production. This incompatibility is more pronounced when the native producer is taxonomically distant from the host organism. It can mainly be related to a difference in codon usage slowing down or ending the mRNA translation. As mentioned above, codons of the foreign sequence can be optimized by replacing rare codons by more common ones for the host [228]. This approach also enabled, in *Aspergillus* species, a stabilization of the transcripts of heterologous genes by preventing premature polyadenylation and subsequent degradation of mRNA [306]. However, this methodology is time consuming in the design of the optimized sequence and relies on still expensive DNA synthesis technology.

At the protein level, the large size of NRPSs increases the possibility to display recognition sequences for host proteases. In particular, the amino acid sequences of inter-module sequence regions, which are largely unfolded (i.e., limited number of α -helix and β -sheet) and display a high degree of flexibility, are therefore more vulnerable to protease activity [307,308]. Yeasts have been described to express multiple vacuolar proteases, such as endoproteinases A and B, potentially released during cell lysis. To prevent the degradation of the newly-synthesized NRPS, a *S. cerevisiae* strain has been mutated in *pep4* and *prb1* genes encoding the previously mentioned proteases [309,310]. The resulting BJ5464 strain has been widely used for heterologous production of various NRPs [57,269,270,272].

In addition, non-native large multi-modular complexes may undergo a complete or partial mis-folding, leading to a loss of enzymatic activity due to rapid degradation of the complex or an inappropriate tertiary structure of catalytic sites. This can be solved or attenuated by co-expressing foreign chaperone encoding genes, which are likely to be involved in the folding of the synthetase in the native host [228]. Alternatively, it was reported that a lower growth temperature could promote the folding of heterologously produced proteins. For example, an increase in the stability of pcbAB synthetase was demonstrated in *H. polymorpha* grown at 25°C instead of 37°C [265]. Nonetheless, this effect was not observed in *S. cerevisiae* with the same NRPS [311]. Moreover, in the enniatin heterologous producer *A. niger*, the effect of the temperature was also investigated among other parameters. It appears that reducing the temperature from 30°C to 26°C did not significantly improve the productivity, especially compared to the impact of the culture medium composition [141].

5.3.4.2. Required enzymes for the biosynthesis of active nonribosomal peptides

In most cases, obtaining the properly folded synthetase is not sufficient to enable the complex to synthesize the compound of interest. Indeed, to be active, NRPSs have to be converted from their *apo* into their *holo* forms through the addition of a flexible phosphopantetheine arm on each thiolation domain by a phosphopantetheinyl transferase. This cofactor is crucial because it carries the monomers previously specifically recognized by the different A-domains. In contrast to filamentous fungi, yeasts (except *S. pombe*) lack a functional PPTase [23]. An endogenous *lys5* PPTase gene, involved in lysine biosynthesis, is expressed in several yeasts but the corresponding enzyme was demonstrated to be unadapted for NRPS activation [222,312]. To complement the lack of a functional PPTase, *sfp* from *B. subtilis* or *npgA* from *A. nidulans*, are generally co-expressed with NRPS genes. NpgA and Sfp have been reported to display a very broad specificity on various NRPSs in diverse fungal and bacterial organisms [43,268,313,314]. The efficiency of these two enzymes was demonstrated in *S. cerevisiae*, and the same order of magnitude of NRP titers was obtained with both of them [311]. Conversely, in *E. coli*, the productivity in terms of cyclodepsipeptides could be increased by 4.8- and 14-fold with *npgA* gene instead of *sfp* and endogenous PPTase, respectively [296]. The *npgA* PPTase gene has even been integrated in the genome of a *S. cerevisiae* strain to generate a cellular platform ready to produce NRP compounds [57,269,270,272].

NRPs synthesized by functional synthetases are not necessarily active compounds, and are often only early intermediates of the targeted bioactive product. Therefore, to obtain the final functional product, addition of gene(s) encoding the tailoring enzyme(s) responsible for NRP modification step(s) is required. This is the case for example, for ACV, which is sequentially modified by three tailoring enzymes to give isopenicillin G (**Figure 18B**).

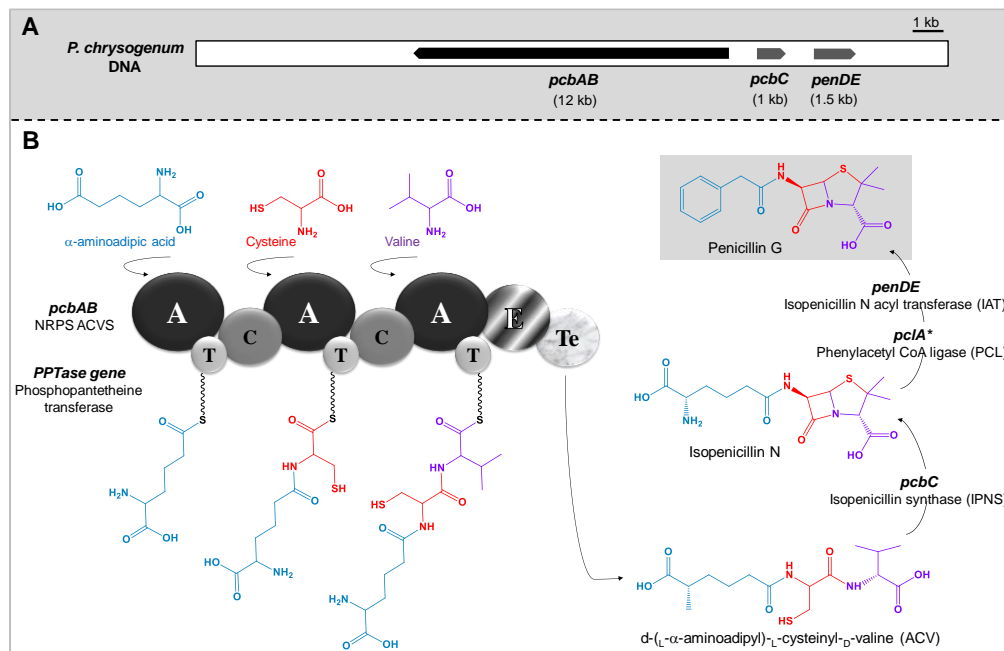


Figure 18: A. Presentation of the gene cluster involved in the biosynthesis of penicillin in *Penicillium chrysogenum*. B. Schematic representation of the penicillin G biosynthetic pathway in *P. chrysogenum*. ACV is assembled by the ACVS (encoded by the *pcbAB* gene), before to be modified into isopenicillin N by IPNS (encoded by *pcbC* gene), subsequently converted into penicillin G through combined activities of IAT and PCL (respectively encoded by *penDE* and *pclA* gene). *PclA** gene is not part of the penicillin biosynthetic cluster.

To heterologously produce isopenicillin G in yeasts, and not only ACV, addition of the isopenicillin N synthase, isopenicillin N acyl transferase and phenylacetyl CoA ligase genes was required [222,265,315]. Moreover, the two last above mentioned tailoring enzymes are located in the peroxisome in the native producer *Penicillium chrysogenum*. This parameter can be crucial, as the efficiency of these enzymes is impacted by their location and environment. An absence of compartmentalization of these enzymes in peroxisomes decreased the productivity by more than 60% in heterologous hosts [222]. This trend was confirmed later in *S. cerevisiae* with an increased productivity gained by tagging these enzymes with PTS1 peroxisome addressing signal [315].

Gliotoxin produced by *A. fumigatus*, is also a bioactive product resulting from multiple modifications of a NRP diketopiperazine encoded by *gliP* synthetase gene. The first step of the biosynthetic pathway from the NRP to gliotoxin was elucidated through a heterologous expression in *A. nidulans* [284,316]. Two tailoring enzymes candidates, namely *gliC* and *gliF* were co-expressed separately with the *gliP* synthetase gene. The disappearance of the precursor NRP in favour of an L-Phe hydroxylated derivative, was only observed in the strain harbouring the cytochrome

P450 monooxygenase encoded by *gliC*. This result demonstrated the role of this tailoring enzyme in the first step of gliotoxin biosynthesis.

Similarly, brevianamide F nonribosomal compound, assembled by FtmA (or FtmS) in *A. fumigatus* and *N. fischeri*, is a platform molecule, precursor to multiple compounds. Wunsch and co-workers have co-transformed *ftmS* in *A. nidulans* with three different prenyltransferase genes from the two previously mentioned fungi. The resulting strains were able to produce, in addition to brevianamide F, up to three supplementary derivative compounds, such as deoxybrevianamide E [280]. With co-expression of *ftmA* with *ftmB*, *ftmC*, *ftmD*, *ftmE*, *ftmF* tailoring enzyme-encoding genes, production of derivatives of brevianamide F was also observed. The NRPS-encoding gene and five tailoring enzyme genes were assembled successively in the same single plasmid, through multiple homologous recombinations in *S. cerevisiae*. All ORFs were positioned in the plasmid under the control of distinct promoters and terminators. After transformation in both *S. cerevisiae* and *A. nidulans*, the enzymatic activity of tailoring enzymes enabled the synthesis of multiple precursors of spirotryprostatin starting from brevianamide F: tryprostatin B, tryprostatin A, and fumitremorgicin [161].

5.3.4.3. How to address NRP precursors deficiency

In the classical ribosomal pathway, only 20 proteinogenic α -amino acid residues are incorporated into proteins. In addition to the proteinogenic α amino acid residues, up to 500 distinct monomers were reported to constitute NRPs building blocks [72]. Among this tremendous biodiversity of monomers, some are directly recognized and incorporated as such, while others are modified during the peptide synthesis by the activity of a NRPS domain. This incredible variety of precursor monomers is the root of the wide diversity of structure and activity of NRPs [31,38]. On the other hand, this feature raises a serious issue regarding heterologous production, as hosts lack the ability to synthesize most of these non-common monomers. A possible approach to complement monomers deficiency relies on the cloning of the genes involved in the monomer biosynthesis pathway. The monomer D-hydroxyisovaleric acid (D-Hiv) is for instance not synthesized by *A. niger* and poorly produced by *S. cerevisiae*. D-Hiv is a precursor engaged in the biosynthesis of cyclodepsipeptides, such as enniatin, beauvericin and bassianolide. It can be obtained by the modification of α -ketovaleric acid by α -ketoisovalerate reductase encoded by *kivR* gene in *B. bassiana* or *F. oxysporum*, among others [141,272]. A two-fold increase of productivity was obtained with the expression of *kivR* gene in a *S. cerevisiae* strain heterologously producing beauvericin (61.7 and 33.8 mg.l⁻¹ with or without *kivR*, respectively) [272]. In *A. niger*, the enniatin heterologous production has also been made possible with the co-expression of the missing *kivR* encoding α -ketoisovalerate reductase with *esyn1* NRPS gene [141].

On the other hand, a precursor deficiency can also be solved by feeding the corresponding compound directly into the culture medium. This approach avoids time consuming gene cloning steps but relies on the host strain being able to take up the

precursor from the medium. The complementation of the above mentioned *A. niger*, expressing foreign enniatin synthetase, was also achieved by feeding it with D-Hiv instead of co-expressing *kivR*. The feeding experiment was quite successful, as it yielded a high enniatin B titer of 950 mg.l⁻¹ after optimization of feeding conditions and culture medium composition. Under the same conditions, the strain fed with 20 mM D-Hiv was produced 25% more than the strain transformed with *kivR* [141]. Nonetheless, the cloning approach remains more cost effective due to the high cost of D-Hiv [272]. Moreover, some adenylation domains display a broader specificity that can be leveraged to generate alternative final products by feeding with alternative precursors. For instance, enniatin B is preferably produced by feeding L-valine in the medium, whereas the alternative enniatin A production is enhanced by feeding with L-leucine or L-isoleucine [141]. In another filamentous fungus, *A. nidulans*, asperphenamate production was knocked down by mutating the *apmA* gene responsible for N-benzoylphenylalaninol biosynthesis. This production was subsequently restored by feeding the same strain with the deficient precursor [283]. In yeast, the relevance of precursor supplementation in the culture medium has been highlighted in *H. polymorpha* expressing *pcbAB* NRPS gene. The production of ACV was not possible due to the incapacity of *H. polymorpha* to synthesize the precursor amino adipic acid. Supplementation of the culture medium with this monomer enabled to unlock the production of ACV [222]. Nonetheless, the use of a precursor feeding strategy is not always a guarantee of success as some negative feedback mechanisms can occur and hinder the compound production. For instance, the feeding of the NRP-precursor 2-aminobutyric acid in the culture medium of *C. asteris*, was recently demonstrated to negatively impact the production of the astin C compound as well as the development of the fungus [317].

Furthermore, the highly developed secondary metabolism of most of the filamentous fungi can hogged available nutrients and precursors. This can lead to a decrease in availability of some monomers and even become a limiting factor for the production of heterologous NRPs. The modification of the metabolic fluxes and the deletion of SM biosynthetic pathways consuming large amount of cell nutrients and energy can be favourable for the production of the foreign compound of interest [155]. Moreover, reducing the secondary metabolite metabolisms of the selected host can facilitate the detection and the purification of the targeted compound by decreasing the presence of unwanted products [231]. Multiple methodologies have been developed to knock out entire SM gene clusters in *A. nidulans*. Among the 71 BGCs predicted in *A. nidulans* [318], the clusters related to the biosynthesis of two major native metabolites of this fungi, the emericellamide and the sterigmatocystin, have been deleted to generate a “cleaner” expression host for PKS/NRPS heterologous production [125].

5.3.4.4. Heterologous NRP toxicity mitigation

A further challenge facing NRP heterologous production is related to the toxicity of some compounds towards the selected host. Native producers have developed various strategies to prevent this toxicity, such as specific secretion mechanisms, compartmentalization of the toxic products, or expression of genes encoding enzymes able to induce a resistance [319–321]. These mechanisms are often specific to the toxic NRP product and therefore may be only present in the native producer and not in the chosen host [322]. The most straightforward strategy to overcome this issue consists to express the NRPS gene under the control of an inducible promoter. This approach enables the choice of the period of production of the toxic compound and to dissociate it from the cell growth period. Hence, GAL1-10, amyB and Tet-on inducible promoters were used to express NRPS genes, respectively in *S. cerevisiae* [311], *A. nidulans* [323], and *A. niger* [141].

Another common approach consists of transferring into the host not only the NRP biosynthesis related genes, but also the specific resistance genes pre-identified into the genome of the native producer. The mechanisms of self-resistance to a produced toxic compound can be classified into three categories of mode-of-operation: (1) by inducing the compound efflux, (2) through enzymatic deactivation of the toxic product, or (3) with a targeted alteration of the drug receptors [324]. Considering that the latter are quite uncommon, and that the deactivation of the toxic product can also impair the targeted activity of the compound, the implementation of an efflux system is often the most straightforward approach to confer a resistance [325]. NRPs are likely to be small structures and accordingly possibly naturally secreted by some organisms. Nonetheless, numerous studies reported a poor secretion of NRPs by heterologous strains such as *H. polymorpha* [222], *S. cerevisiae* [315], and *A. niger* [295]. Therefore, expression of a foreign NRP specific transport protein can be required to improve the secretion procedure [326]. However, the identification of resistance genes (regardless of their mode-of-operation) in the native producer is not always easy. In filamentous fungi, genes conferring resistance to toxic secondary metabolites have been highlighted multiple time within the corresponding gene cluster. For instance, in the gliotoxin BGC of *A. fumigatus*, the *gliA* and *gliT* genes respectively encode an efflux pump and a gliotoxin oxidoreductase which can reduce the toxicity of the compound [327,328]. Nonetheless, these resistance genes are not systematically located within the corresponding gene cluster, and the annotation methods are often far to be sufficiently efficient to identify precisely resistance genes, especially in fungal organisms [329].

There are further options to protect the heterologous host from the toxicity of the foreign compound, such as extracting the product from the culture medium. The toxic compound can for instance be extracted by introducing a layer of organic solvent on the top of the culture, where the product can diffuse [319,330]. Some absorber resins are also able to bind the toxic compound when they are added to the culture broth. This was for instance demonstrated with the Diaion HP20 resin binding the mycotoxin cercosporamide contained in a fungal fermentation broth, thus enabling to increase by 100-fold the final product yield [331].

5.3.4.5. From the heterologous host to the *in vitro* production assays: a viable approach?

5.3.4.5.1. Approach and methodologies

The *in vitro* NRP synthesis constitutes a cell-free production system where the enzymes involved in the biosynthetic pathway (NRPS gene, tailoring enzymes) are purified and subsequently incubated with the precursor monomers to generate the desired compound. The *in vitro* approach is a suitable solution to address most of the issues described above. Indeed, it addresses the deficiency in NRP precursors, but also circumvents issues of NRP toxicity towards the host strain. In order to perform *in vitro* assays, the synthetase and/or associated tailoring enzymes need to be purified in soluble form and in sufficient quantity from the heterologous host. Addition of a tag sequence, either on the C-terminal or on the N-terminal part of the heterologously expressed synthetase, is highly recommended to facilitate further extraction-purification steps. The use of 6xHis tag [269,270,282] or Flag tag [57], respectively purified through metal or antibody affinity chromatography, have been reported in previous *in vitro* NRPS studies. Alternatively, NRPSs were also purified by size exclusion chromatography as their high molecular weight facilitates their separation from other host proteins [84,332].

Thereafter, it is possible to assess the presence of the phosphopantetheine arm on thiolation domains through three approaches. The most recently developed method consists to generate a peptide mixture by a trypsin digestion of the proteins contained in the cell extract, before to separate fractions from the peptide mixture through a strong cation exchange chromatography. Subsequently, a MS/MS analysis enables to detect into the different fractions the presence of unique marker ions derived from the NRPS-bonded phosphopantetheine arm [333]. Alternatively, the presence of the phosphopantetheine arm on the *holo*-NRPS can also be evaluated through the thioester-binding assay measuring the consumption of radiolabeled [¹⁴C] amino-acids incorporated in the synthesized NRP [334]. An older technic consists to perform an alkaline phosphatase treatment to hydrolyze the phosphopantetheine part from the *holo*-NRPS. This treatment releases a pantothenic acid which can be quantified through a microbiological assay against *Lactobacillus plantarum* [84]. Regarding to A-domains, the most common approach to confirm their enzymatic activity consists to quantify the incorporation of ³²Ppi into ATP in presence of the precursor amino acids (ATP pyro-phosphate exchange radioassay) [282,332]. Alternatively, a nonradioactive method has also been developed enabling to evaluate the A-domains activity through a MS/MS analysis. This technic relies on the measurement of the isotopic back exchange of the pyrophosphate into a [¹⁸O]-labeled ATP [335]. Recently, Duckworth and coworkers (2016) have developed another method based on the measurement of the pyrophosphate released during the adenylation reaction. The pyrophosphate released subsequently to the action of an active A-domain is first cleaved into an inorganic phosphate by an inorganic pyrophosphatase. Subsequently, a purine nucleoside phosphorylase uses this inorganic pyrophosphate in a reaction converting 7-methylthioguanosine into 7-methylthioguanine which can be monitored

with a spectrometric-based method at 360 nm [336]. Finally, the functionality of the *holo*-NRPS complex can be confirmed with the identification of the final NRP by using appropriate analytical methods (LC-MS and/or NMR).

5.3.4.5.2. Successful applications related to fungal NRPs

Genome mining and *in silico* approaches have enabled, in *C. subvermispora*, to assign the basidioferrin siderophore synthesis to the *CsNPS2* NRPS-encoding gene. Experimental validation of these predictions has been performed through heterologous production coupled with *in vitro* assays. *CsNPS2* was expressed in *A. niger* and subsequently purified through metal affinity chromatography. The single A-domain of *CsNPS2* displayed specificity for two substrates: N₅-acetyl-N₅-hydroxy-L-ornithine (L-AHO) and N₅-cis-anhydromevalonyl-N₅-hydroxy-L-ornithine (L-AMHO). Nonetheless, basidioferrin was only detected after 24 h of *in vitro* assay with 1 mM L-AHO and not with L-AMHO. These results confirmed the role of *CsNPS2* in the basidioferrin biosynthesis [282].

Several NRPSs were also heterologously expressed and purified from *S. cerevisiae* for the purpose of *in vitro* assays: AnaPS [57], TqaA [57], AspA [270], ArdA [269]. The presence of the *npgA* gene in the genome of the yeast was compulsory to produce the PPTase enabling to convert the synthetase from the *apo* to the *holo* activated form. *In vitro* assays with 10 μM pure AspA with its monomer Ant (1 mM) and L-Trp (1 mM) generated a tripeptide, the iterative mode-of-operation of this bimodular NRPS [270]. The TqaA synthetase is to date the largest NRPS protein (450 kDa) whose activity has been reconstituted *in vitro* in its active form [57]. The ArdA synthetase from *N. fischeri*, displaying high homology with TqaA, was also used to perform *in vitro* assays after being purified from a heterologous *S. cerevisiae* producer. This 430-kDa bimodular NRPS was demonstrated *in vitro* to sufficiently assemble Ant, L-Ala, and L-Trp into ardeemin FQ [269].

In another *in vitro* study aiming at obtaining NRP derivatives, Tsunematsu *et al.* (2013) first produced and purified the different intermediates of the spirotryprostatin biosynthetic pathways in a heterologous *S. cerevisiae*. In the meantime, tailoring enzymes FtmG and FqzB were also purified from heterologous yeast. With the purpose of deciphering the role of these tailoring enzymes in the final step of spirotryprostatin synthesis, *in vitro* assays were performed with the purified precursors mentioned above. As a result of this work, *fimG* encoding a cytochrome P450 has been shown to be involved in spiro-ring formation in spirotryprostatin B, while the FAD-dependent monooxygenase encoded by *fqzB* was shown to be responsible for the spiro-carbon formation in spirotryprostatin A [161].

The *in vitro* NRP assembling strategy is especially useful as a complement to heterologous production by facilitating the elucidation of synthetase mode-of-operation or tailoring enzymes function. Nonetheless, due to the very low amounts of NRPS obtained through affinity chromatography purification, this approach is not promising for increasing NRP yields. Moreover, the cost incurred by the addition of

precursor monomers and ATP largely prevents its applications for large-scale NRP production.

5.4. *Penicillin heterologous production: a NRP model system*

Due to the clinical and commercial importance of β -lactam antibiotics, researchers have attempted for years to develop their heterologous production in hosts renowned for being readily cultivate at industrial scale, and easy to handle through genetic engineering. This multi-decade work has positioned the NRP-derived penicillin compounds as a model system regarding NRP heterologous production. These efforts have enabled both to better understand the ins and outs of the penicillin biosynthetic pathway and to develop various strategies to implement heterologous production in fungal cell factories.

Penicillin β -lactam antibiotics are naturally produced by some *Penicillium* species, especially *P. chrysogenum* and to a lesser extent by some *Aspergillus* species such as *A. nidulans* [113,337]. In *P. chrysogenum*, the elucidated penicillin biosynthetic pathway relies on the joint action of several enzymes, whose corresponding genes are mainly located within the same gene cluster (**Figure 18A**). The *pcbAB* gene encodes a 3-module NRPS that is responsible for assembling ACV (**Figure 18B**). This nonribosomal peptide precursor of penicillin, is then modified into isopenicillin (IPN) by an isopenicillin N synthase. This intermediate is further converted by an isopenicillin N acyl transferase (IAT) and a phenylacetyl CoA ligase (PCL) into penicillin G [222]. The penicillin V in which the benzyl group is replaced by a phenoxyethyl group is a commercial analog of the penicillin G displaying an enhanced stability in acidic environment. This analog is preferably produced by *P. chrysogenum* when phenoxyacetic acid is added in the fermentation medium [338]. This appears reasonable because the production of low amounts of penicillin compounds, and the presence of enzymes sharing a high level of sequence identity with those from *P. chrysogenum*, had been reported in [337]. An associated positive impact of these native enzymes cannot be ruled out to explain the success of this heterologous production.

The first transfer of the penicillin biosynthetic pathway in a heterologous host was achieved into *N. crassa* and *A. niger*. The cosmid pCX3.2 containing the ACVS encoding gene cluster was cloned into both heterologous hosts, leading directly to production of penicillin. The presence of the produced antibiotic was successfully assessed through a bioactivity assay against *B. subtilis*. *A. niger* was reported to produce much more penicillin V than *N. crassa* (up to 2.3 $\mu\text{g}\cdot\text{ml}^{-1}$ versus 0.170 $\mu\text{g}\cdot\text{ml}^{-1}$) [278]. One explanation could be the lack of ACV precursor availability in *N. crassa*. A second hypothesis rests on the fact that *A. niger* and *P. chrysogenum* are closely related species both belonging to the order Eurotiales, unlike *N. crassa* which is associated to Sordariales.

A few years later, the penicillin biosynthetic pathway was also reconstituted into the methylotrophic yeast *H. polymorpha*, which is well-known for its superior fermentation properties even at industrial scale [339]. The 12-kb *pcbAB* NRPS gene was inserted, under the control of the strong methanol-inducible promoter from alcohol oxidase gene (P_{AOX}) in pHIPZ4 plasmid. *H. polymorpha* was transformed with the linearized plasmid in order to integrate the gene at the AOX locus. Subsequently,

heterologous expression of ACV synthetase (ACVS) was assessed directly at the protein level by Western blot with antibodies directed to ACVS amino-acid sequence. As described in the section 3, the homologous PPTase Lys5 of the methylotrophic yeast was not adapted to activate ACVS, and an additional *sfp* PPTase encoding-gene was inserted. Subsequently, IPNS, IAT and PCL-encoding genes from *P. chrysogenum* were integrated into the genome step by step. Penicillin G productivity of the resulting yeast was similar to those obtained in the native producer NRRL1951 (1.1 and 1.0 $\mu\text{g}\cdot\text{ml}^{-1}$, respectively). In the same yeast, deletion of the native *PEX3* gene, responsible for the peroxisome proliferation, led to a 64% reduced yield (0.4 $\mu\text{g}\cdot\text{ml}^{-1}$) [222]. The importance of the peroxisomal location of IAT and PCL had previously been highlighted in *P. chrysogenum* and *A. nidulans* [340].

In subsequent work, *S. cerevisiae* was also selected to heterologously produce ACV [311]. *pcbAB* was inserted into a replicative multi-copy pESC by TAR cloning under the control of the galactose inducible promoter GAL1/GAL10 ($P_{\text{GAL1/GAL10}}$). The resulting construct bearing an *URA3* selection gene was used to transform a *S. cerevisiae* Δura3 strain. One of the assets of pESC is the presence of the bi-directional $P_{\text{GAL1/GAL10}}$ enabling the cloning of a second gene. This allows the cloning of the PPTase encoding gene in the same vector as the NRPS gene (if the NRPS is encoded by a single ORF). Under galactose induction, the transformed strain carrying this plasmid was able to express functional ACVS (transcription assessed by qPCR and detection of the NRP by LC-MS analysis). A codon optimization of the 5' portion of the ACVS gene was also performed leading to a slight, but not significant, increase of the productivity. On the other hand, *pcbAB* was also integrated into a long terminal repeats (LTR) retrotransposons locus of *S. cerevisiae*. LTR retrotransposons are transposable elements present in high number and spread all over the yeast genome, thus making these sites suitable for multi-copy genomic integration [341]. Yields obtained with genomic integration were 6-fold lower than in the multi-copy PESC plasmid, demonstrating the benefits of the latter expression system. A few years later, another ACV producing *S. cerevisiae* was obtained with the cloning of *pcbAB* and *npgA* genes into a low-copy centromeric plasmid [315]. Multiple pairs of promoters were tested to enhance the expression of both the *pcbAB* and *npgA* genes. The combination of *pcbAB* and *npgA*, respectively under the control of P_{TDH3} and P_{PGK1} constitutive promoters, led to a 4-fold higher ACV productivity than with $P_{\text{GAL1/GAL10}}$ in PESC plasmid (≈ 0.28 versus ≈ 0.07 $\mu\text{g}\cdot\text{ml}^{-1}$). This demonstrated the importance of promoter choice for each gene to be expressed, and the underestimated requirement for compatibility of each promoter/gene pair. Subsequently, the complete biosynthetic pathway to the penicillin G was reconstituted by addition of tailoring enzymes IPN, IAT and PCL. Expression of the tailoring enzymes were also evaluated under multiple combinations of promoters for the sake of optimizing the production. The resulting strain was able to efficiently produce and secrete penicillin G, but with yields still two fold less than in heterologous *H. polymorpha* [222].

With the purpose of transferring the multiple genes involved in the penicillin synthesis from *P. chrysogenum* into *A. nidulans* SAA248 lacking a native penicillin gene cluster, Unkles and coworkers have exploited the previously described properties

of the viral 2A peptide. By inserting viral 2A peptide between each single ORF, it is possible to express from the same plasmid a polycistronic mRNA under the control of a single strong/inducible promoter [258]. Viral 2A sequences were inserted between *pcbAB*, IPNS gene (*pcbC*), and IAT gene (*penDE*). The resulting 15.5-kb sequence was cloned by *in vivo* homologous recombination in *S. cerevisiae*, into pJET1.2 plasmid under the control of xylose inducible promoter. The obtained plasmid was then transferred to *A. nidulans*, leading to the transcription of the 15.5-kb polycistronic mRNA. Subsequently, each individual enzyme and especially the massive NRPS were cleaved and their joint action enabled the synthesis of penicillin G. Notably, the 331-T25 mutant having integrated multiple copies of the genes, was able to produce 3-fold more penicillin G than the wild type *A. nidulans* strain 0760 (0.687 versus 0.234 $\mu\text{g}\cdot\text{ml}^{-1}$) [226]. Interestingly, in addition to the heterologous IAT enzyme, a second native IAT was identified in the transformed SAA248 strain. Even if *A. nidulans* SAA248 has lost its ability to synthesize penicillin like other *A. nidulans* strains, the presence of genes encoding enzymes involved in the penicillin biosynthetic pathway seems to persist in its genome [337]. An associated positive impact of these native enzymes cannot be ruled out to explain the success of this heterologous production.

It is not straightforward to compare the penicillin production yields obtained with the different heterologous strains described in this section. Nonetheless, authors have each time compared these yields with those observed either in the *P. chrysogenum* NRRL1951 or *A. nidulans* 0760 native penicillin producers, enabling some cross-study comparisons. Regarding to yeast heterologous systems, *H. polymorpha* was reported to produce similar penicillin G amount to *P. chrysogenum* NRRL1951 strain ($\sim 1 \mu\text{g}\cdot\text{ml}^{-1}$) [222]. In contrast, *S. cerevisiae* was half efficient, probably due to the considerably less developed peroxisomal system compared to the methylotrophic yeast [315]. Regarding to filamentous fungi heterologous systems, the 331-T25 mutant of *A. nidulans* SAA248 expressing penicillin biosynthetic genes coming from *P. chrysogenum* was 3-fold more efficient than the native penicillin G producer *A. nidulans* 0760 ($\sim 0.69 \mu\text{g}\cdot\text{ml}^{-1}$), but still 30% less productive than *P. chrysogenum* NRRL1951 [226]. In terms of heterologous penicillin V production, wide variations have been observed between the two described hosts with up to 13-fold higher productivity in *A. niger* compared to *N. crassa*. The highest penicillin V concentrations reported in the heterologous *A. niger* strain was likely due to the close taxonomic relatedness of this strain with *P. chrysogenum* [278].

Over many years, the heterologous production of penicillin has been a long-standing test case of study. It enabled researchers to deal with most of the key factors limiting NRP heterologous production and to unlock some of them. Unfortunately, in terms of productivity, the heterologous production approach is decades behind on current industrial penicillin G production systems. Indeed, several studies were conducted to generate and screen thousands of *P. chrysogenum* mutants, with the aim of improving production yields. Industrial production, with these selected strains, has been optimized for years in terms of strain productivity and fermentation process [342,343]. The analysis of the genome of the strains with an improved penicillin production

revealed the presence of multiple copies of the penicillin biosynthetic genes, which is likely to contribute in the increased productivity [113,344,345]. Another characteristic of high-penicillin producing strains is the loss of several gene clusters involved in secondary metabolite synthesis which were detected in the wild-type *P. chrysogenum* NRRL195. For instance, the improved penicillin producer *P. chrysogenum* Wisconsin 54-1255 lacks the ability of producing sorbicillin and penitric acid which are present in the genome of the parental NRRL195 strain. The deletion of secondary metabolite biosynthetic pathways also contributed to improve the strain productivity for β -lactam antibiotics [346,347]. Although the current penicillin G industrial production yields are not available in the literature, these yields were reported few years ago to be already more than 1000-fold higher than those obtained with the wild-type *P. chrysogenum* NRRL195 [348].

5.5. Concluding remarks

The recent development of high-throughput sequencing technologies combined with substantial drop in sequencing costs have driven an impressive increase in the number of available fungal genomes. This factor coupled with the improvement of genome annotating methodologies has partially unlocked the SM production potential of fungi. Hence, efficient bioinformatics tools now enable a more accurate analysis in terms of nonribosomal pathway identification and substrate specificity predictions. Nonetheless, the prediction robustness is lowered because already elucidated fungal NRPs and NRPSs remain under-represented in dedicated databases as compared to bacterial ones. Consequently, the predictions based on genomic sequences need to be validated by structural elucidation of the NRPs which implies its efficient production.

The production without genetic manipulation of NRP compounds identified *in silico* is not systematically possible in native hosts due to the existence of cryptic biosynthetic gene clusters and/or incompatibility between the strain and the laboratory cultivation conditions. Therefore, the heterologous production approach is often the most appropriate strategy to deal with this issue. Multiple fungal cellular platforms have been developed to heterologously produce foreign fungal nonribosomal peptides in particular. Among others, the yeast *S. cerevisiae* and filamentous fungi related to *Aspergillus* spp. have been widely used. Heterologous production of NRPs is particularly challenging due to a plethora of potential limiting steps. Combining multiple strategies to optimize each of them can greatly increase the chances of successful production. Especially, the transfer of large NRPS sequences encompassing tens of kb is one of the main hurdles to overcome. Several methodologies, such as TAR cloning, ExRec, Gibson assembly, USER cloning, and overlap extension PCR, have been developed to enable the reconstruction of single large NRPS genes or gene clusters. This laborious and time-consuming engineering of large NRPS biosynthetic pathways can be avoided by *de novo* DNA synthesis, but this remains prohibitively costly in many cases. However, considering that costs of gene synthesis are continuously decreasing, an intensification of the use of this rapid methodology for transferring foreign NRPS genes and gene clusters can be expected in the upcoming years.

The correct transfer of the targeted gene in the heterologous host is not a guarantee of success for the production of the related NRP. For instance, Harvey and coworkers (2018) have transferred 41 cryptic fungal gene clusters to *S. cerevisiae* but only 54% have been processed by the yeast to lead to a detectable compound [349]. This result, while promising, demonstrates the challenges related to the heterologous expression of large fungal gene clusters. Nonetheless, the efficiency of the heterologous production of NRPs can be greatly improved by taking into account some key factors. First, the transcription rate can be improved by using an appropriate promoter, and sometimes with adequate initiation factors. Translation and proper folding of the protein are also crucial factors to consider. Optimization of this can be achieved by expression of foreign chaperones enzymes or with different cultivation conditions. It is also important to protect the foreign NRPS from the proteolytic activity of the

selected host. Once the properly folded NRPS is obtained and activated via a homologous or heterologous phosphopantetheinyl transferase, the assembly of the NRP can be considered. At this last stage, NRP precursor deficiency or NRP toxicity against the selected hosts can also become limiting factors. The first one can be overcome by feeding or cloning of the missing genes related to precursor biosynthesis, while the second may be prevented by conferring secretion ability to the host with the cloning of transporter-associated genes, for example.

Despite these drawbacks, heterologous production strategies were successfully employed to identify novel NRP compounds, to decipher NRPS modes-of-operation and in some cases, to improve productivity. This approach is especially relevant to producing compounds for which the corresponding genes are cryptic or weakly expressed in native organisms for which no genetic tools are available yet. Conversely, transferring a NRP biosynthetic pathway to a heterologous host is often not sufficient to overcome production yield obtained with a native producer already optimized and well industrially established strain as in the penicillin case. Overall, the NRP production rates currently obtained with heterologous expression do not yet reach levels suitable for industrial production. Chemical synthesis remains in most of the cases, the more competitive way to obtain peptides and especially regarding linear peptides including only proteinogenic amino acids. A possible exception concerns certain high-value compounds, which are cyclic, or include unusual building block and/or monomers which are further modified by tailoring enzyme. Indeed, these specific characteristics are often difficult to obtain via chemical synthesis and should be obtained more easily through a heterologous expression of the corresponding NRPS-encoding gene.

- **Conflict of interest**

Authors of the manuscript declare there is no conflict of interests.

- **Aknowledgements**

This work was supported by the CPER Alibiotech project; the INTERREG V France-Wallonie-Vlaanderen Project SmartBioControl/BioScreen; the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (FRIA); and the international laboratory Microbial Secondary Metabolites MOM (University of Lille/Gembloux Agro-Bio Tech University of Liege). Grant PhD co-funding of the University of Liege (Welcome Grant Philippe Jacques) and the University of Lille.

6. References

1. Bennett, J.W., and Bentley, R. (1989) What's in a Name?-Microbial Secondary Metabolism. *Adv. Appl. Microbiol.*, **34** (C), 1–28.
2. Sertuerner (1817) Ueber das Morphium, eine neue falzfahige Grundlage, und die Mekonfaure, als Hauptbestandtheile des Opiums. *Pharmac. zu Eimbeck im Konigr. Hann.*, 56–89.
3. Raistrick, H. (1950) A region of Biosynthesis. *London Sch. Hyg. Trop. Med.*, **136** (January), 481–509.
4. Fleming, A. (1929) On the antibacterial action of cultures of a *penicillium*, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.*, **X** (3), 226–236.
5. Dias, D.A., et al. (2012) A Historical Overview of Natural Products in Drug Discovery. *Metabolites*, **2** (2), 303–336.
6. Baxter, A., et al. (1992) Squalenstatin 1, a potent inhibitor of squalene synthase, which lowers serum cholesterol in vivo. *J. Biol. Chem.*, **267** (17), 11705–11708.
7. Weber, G., et al. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.*, **26** (2), 120–125.
8. Finkelstein, E., et al. (1996) Griseofulvin and its uses. *Int. J. Antimicrob. Agents*, **6** (4), 189–194.
9. Udagawa, T., et al. (2000) Cytochalasin E, an Epoxide Containing *Aspergillus*-Derived Fungal Metabolite, Inhibits Angiogenesis and Tumor Growth. *J. Pharmacol. Exp. Ther.*, **294** (2), 421–427.
10. Stassen, P.M., et al. (2007) Use of mycophenolic acid in non-transplant renal diseases. *Nephrol. Dial. Transplant.*, **22** (4), 1013–1019.
11. Alberti, F., et al. (2017) Natural products from filamentous fungi and production by heterologous expression. *Appl. Microbiol. Biotechnol.*, **101** (2), 493–500.
12. Campbell, I.M. (1985) Secondary Metabolism and Microbial Physiology. *Adv. Microb. Physiol.*, **25** (C), 1–60.
13. Keller, N.P., et al. (2005) Fungal secondary metabolism - From biochemistry to genomics. *Nat. Rev. Microbiol.*, **3** (12), 937–947.
14. Berg, T.L., et al. (1965) the Biosynthesis of Gramicidin S in a Cell-Free System. *Biochem. J.*, **96**, 43–52.
15. Lipmann, F., et al. (1971) Polypeptide synthesis on protein templates: the enzymatic synthesis of gramicidin S and tyrocidine. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **35**, 1–34.
16. Schafhauser, T., et al. (2019) Antitumor astins originate from the fungal endophyte *Cyanoderma asteris* living within the medicinal plant *Aster tataricus*. *Submitted*.
17. Kleinkauf, H., and Von Döhren, H. (1996) A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.*, **236** (2), 335–351.
18. Yagasaki, M., and Hashimoto, S.I. (2008) Synthesis and application of dipeptides; Current status and perspectives. *Appl. Microbiol. Biotechnol.*, **81** (1), 13–22.
19. Demain, A.L. (2014) Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.*, **41** (2), 185–201.
20. Wang, H., et al. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci.*, **111** (25), 9259–9264.
21. Calteau, A., et al. (2014) Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics*, **15** (1), 1–14.
22. Bushley, K.E., et al. (2008) Module evolution and substrate specificity of fungal nonribosomal peptide synthetases involved in siderophore biosynthesis. *BMC Evol. Biol.*, **8** (1), 1-24.

23. Schwecke, T., et al. (2006) Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. *ChemBioChem*, **7** (4), 612–622.
24. Mootz, H.D., and Marahiel, M.A. (1997) The Tyrocidine Biosynthesis Operon of. *Microbiology*, **179** (21), 6843–6850.
25. Weber, T. (2014) *In silico* tools for the analysis of antibiotic biosynthetic pathways. *Int. J. Med. Microbiol.*, **304** (3–4), 230–235.
26. Finking, R., and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.*, **58**, 453–488.
27. Imperi, F., and Visca, P. (2013) Subcellular localization of the pyoverdine biogenesis machinery of *Pseudomonas aeruginosa*: A membrane-associated “siderosome.” *FEBS Lett.*, **587** (21), 3387–3391.
28. Hantash, F.M., and Earhart, C.F. (2000) Membrane association of the *Escherichia coli* enterobactin synthase proteins EntB/G, EntE, and EntF. *J. Bacteriol.*, **182** (6), 1768–1773.
29. Mohamed A. Marahiel, et al. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.*, **97** (7), 2651–2674.
30. Walsh, C.T., et al. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.*, **5** (5), 525–534.
31. Schauwecker, U.K. and F. (2003) Combinatorial Biosynthesis of Non-Ribosomal Peptides. *Comb. Chem. High Throughput Screen.*, **6** (6), 527–540.
32. Esmaeel, Q., et al. (2016) Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism catalyzes amonabactin synthesis in *Aeromonas*. *Appl. Microbiol. Biotechnol.*, **100** (19), 8453–8463.
33. Stachelhaus, T., et al. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, **6** (8), 493–505.
34. Conti, E., et al. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.*, **16** (14), 4174–4183.
35. May, J.J., et al. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U. S. A.*, **99** (19), 12120–12125.
36. Miyanaga, A., et al. (2014) The crystal structure of the adenylation enzyme VinN reveals a unique β -amino acid recognition mechanism. *J. Biol. Chem.*, **289** (45), 31448–31457.
37. Niquille, D.L., et al. (2018) Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.*, **10** (3), 282–287.
38. Challis, G.L., and Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Struct. Biol.*, **14** (6), 748–756.
39. Sy, J., et al. (1973) Nonribosomal synthesis of guanosine 5',3'-polyphosphates by the ribosomal wash of stringent *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.*, **70** (7), 2145–2148.
40. Clark, L., et al. (2018) *In silico* analysis of class I adenylate-forming enzymes reveals family and group-specific conservations. *PLoS One*, **13** (9), e0203218.
41. Beld, J., et al. (2014) The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.*, **31** (1), 61–108.
42. Schlumbohm, W., et al. (1991) An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem.*, **266** (34), 23135–23141.
43. Quadri, L.E.N., et al. (1998) Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Society*, **37** (January), 1585–1595.

44. Stachelhaus, T., et al. (1996) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.*, **3** (11), 913–921.
45. Weber, T., et al. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure*, **8** (4), 407–418.
46. Weber, T., and Marahiel, M.A. (2001) Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure*, **9** (1), 3–9.
47. Keating, T.A., et al. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.*, **9** (7), 522–526.
48. Zhang, J., et al. (2016) Structural basis of nonribosomal peptide macrocyclization in fungi. *Nat. Chem. Biol.*, **12** (12), 1001–1003.
49. Belshaw, P.J., et al. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284** (5413), 486–489.
50. De Crécy-Lagard, V., et al. (1995) Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C. R. Acad. Sci. III.*, **318** (9), 927–936.
51. Stachelhaus, T., et al. (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.*, **273** (35), 22773–22781.
52. Kohli, R.M., et al. (2002) The thioesterase domain from a nonribosomal peptide synthetase as a cyclization catalyst for integrin binding peptides. *Proc. Natl. Acad. Sci.*, **99** (3), 1247–1252.
53. Bruner, S.D., et al. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure*, **10** (3), 301–310.
54. Claire C. Tseng, et al. (2002) Characterization of the Surfactin Synthetase C-Terminal Thioesterase Domain as a Cyclic Depsipeptide Synthase. *Biochemistry*, **41** (45), 13350–13359.
55. Koglin, A., and Walsh, C.T. (2009) Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep.*, **26** (8), 987–1000.
56. Manavalan, B., et al. (2010) Molecular modeling of the reductase domain to elucidate the reaction mechanism of reduction of peptidyl thioester into its corresponding alcohol in nonribosomal peptide synthetases. *BMC Struct. Biol.*, **10** (1), 1.
57. Gao, X., et al. (2012) Cyclization of Fungal Nonribosomal Peptides by a Terminal Condensation-Like Domain. *Nat Chem Biol*, **412** (2), 426–434.
58. Samel, S.A., et al. (2014) Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallogr. D. Biol. Crystallogr.*, **70** (Pt 5), 1442–1452.
59. Kim, W.E., et al. (2019) Mechanistic Probes for the Epimerization Domain of Nonribosomal Peptide Synthetases. *ChemBioChem*, **20** (2), 147–152.
60. Linne, U., and Marahiel, M.A. (2000) Control of Directionality in Nonribosomal Peptide Synthesis: Role of the Condensation Domain in Preventing Misinitiation and Timing of Epimerization. *Biochemistry*, **39** (34), 10439–10447.
61. McErlean, M., et al. (2019) Refining and expanding nonribosomal peptide synthetase function and mechanism. *J. Ind. Microbiol. Biotechnol.*, **46** (3–4), 493–513.
62. Reimer, J.M., et al. (2016) Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature*, **529** (7585), 239–242.
63. Reimer, J.M., et al. (2018) Piecing together nonribosomal peptide synthesis. *Curr. Opin. Struct. Biol.*, **49**, 104–113.
64. Winn, M., et al. (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.*, **33** (2), 317–347.
65. Chen, H., et al. (2001) Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem. Biol.*, **8** (9), 899–912.

66. Du, L., et al. (2000) An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. *FEMS Microbiol. Lett.*, **189** (2), 171–175.
67. Schneider, T.L., et al. (2003) Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry*, **42** (32), 9722–9730.
68. Hu, F., et al. (2019) Rational strain improvement for surfactin production: enhancing the yield and generating novel structures. *Microb. Cell Fact.*, **18** (1), 42.
69. Hoyer, K.M., et al. (2007) The Iterative Gramicidin S Thioesterase Catalyzes Peptide Ligation and Cyclization. *Chem. Biol.*, **14** (1), 13–22.
70. Caboche, S., et al. (2008) NORINE: A database of nonribosomal peptides. *Nucleic Acids Res.*, **36** (1), 326–331.
71. Calcott, M.J., and Ackerley, D.F. (2015) Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol.*, **15** (1), 162.
72. Caboche, S., et al. (2010) Diversity of monomers in nonribosomal peptides: Towards the prediction of origin and biological activity. *J. Bacteriol.*, **192** (19), 5143–5150.
73. Lang, C., and Staiger, C. (2016) Tyrothricin--An underrated agent for the treatment of bacterial skin infections and superficial wounds? *Pharmazie*, **71** (6), 299–305.
74. Müller, W.H., et al. (1991) Localization of the pathway of the penicillin biosynthesis in *Penicillium chrysogenum*. *EMBO J.*, **10** (2), 489–495.
75. Nicolaou, K.C., et al. (1999) Chemistry, Biology, and Medicine of the Glycopeptide Antibiotics. *Angew. Chemie Int. Ed.*, **38** (15), 2096–2152.
76. Krause, C., et al. (2006) Sequence diversity of the peptaibol antibiotic suzukacillin-A from the mold *Trichoderma viride*. *J. Pept. Sci.*, **12** (5), 321–327.
77. Hüttel, W., et al. (2016) Echinocandin B biosynthesis: a biosynthetic cluster from *Aspergillus nidulans* NRRL 8112 and reassembly of the subclusters Ecd and Hty from *Aspergillus pachycristatus* NRRL 11440 reveals a single coherent gene cluster. *BMC Genomics*, **17** (1), 570.
78. Dittmann, E., et al. (2001) Molecular biology of peptide and polyketide biosynthesis in cyanobacteria. *Appl. Microbiol. Biotechnol.*, **57** (4), 467–473.
79. Pedras, M.S.C., and Khallaf, I. (2012) Molecular interactions of the phytotoxins destruxin B and sirodesmin PL with crucifers and cereals: Metabolism and elicitation of plant defenses. *Phytochemistry*, **77**, 129–139.
80. Cavelier, F., et al. (1998) Natural cyclopeptides as leads for novel pesticides: tentoxin and destruxin. *Pestic. Sci.*, **52** (1), 81–89.
81. Orlando, B., et al. (2019) *Fusarium* species and enniatin mycotoxins in wheat, durum wheat, triticale and barley harvested in France. *Mycotoxin Res.*, 1-12.
82. Mallebrera, B., et al. (2018) *In vitro* mechanisms of Beauvericin toxicity: A review. *Food Chem. Toxicol.*, **111**, 537–545.
83. Coutte, F., et al. (2010) Production of surfactin and fengycin by *Bacillus subtilis* in a bubbleless membrane bioreactor. *Appl. Microbiol. Biotechnol.*, **87** (2), 499–507.
84. Lawen, A., and Zocher, R. (1990) Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described. *J. Biol. Chem.*, **265** (19), 11355–11360.
85. Zafrir-Ilan, E., and Carmeli, S. (2010) Eight novel serine proteases inhibitors from a water bloom of the cyanobacterium *Microcystis* sp. *Tetrahedron*, **66** (47), 9194–9202.
86. Morita, H., et al. (1995) Solution forms of antitumor cyclic pentapeptides with 3,4-dichlorinated proline residues, astins A and C, from *Aster tataricus*. *Chem. Pharm. Bull. (Tokyo)*, **43** (8), 1395–1397.

87. Maiya, S., et al. (2006) The fumitremorgin gene cluster of *Aspergillus fumigatus*: Identification of a gene encoding brevianamide F synthetase. *ChemBioChem*, **7** (7), 1062–1069.
88. Morita, H., et al. (1993) Astins A and B, antitumor cyclic pentapeptides from *Aster tataricus*. *Chem. Pharm. Bull. (Tokyo)*, **41** (5), 992–993.
89. Jiang, J., et al. (1994) Approaches toward the total synthesis of astins A, B, and C. *Tetrahedron Lett.*, **35** (14), 2121–2124.
90. Wang, L., et al. (2014) Astin B, a cyclic pentapeptide from *Aster tataricus*, induces apoptosis and autophagy in human hepatic L-02 cells. *Chem. Biol. Interact.*, **223**, 1–9.
91. Ludwig-Müller, J., et al. (2014) Improvement of hairy root cultures and plants by changing biosynthetic pathways leading to pharmaceutical metabolites: Strategies and applications. *Biotechnol. Adv.*, **32** (6), 1168–1179.
92. Li, F., et al. (2018) Design and synthesis of plant cyclopeptide Astin C analogues and investigation of their immunosuppressive activity. *Bioorganic Med. Chem. Lett.*, **28** (14), 2523–2527.
93. Morita, H., et al. (1994) Conformational analysis of an antitumour cyclic pentapeptide, astin B, from *Aster tataricus*. *Tetrahedron*, **50** (40), 11613–11622.
94. Cheng, D.L., et al. (1996) New pentapeptides from *Aster tataricus*. *Phytochemistry*, **41** (1), 225–227.
95. Kosemura, S., et al. (1993) Isolation and Structure of Asterin, a new halogenated cyclic pentapeptide from *Aster tataricus*. *Tetrahedron Lett.*, **34** (8), 1291–1294.
96. Morita, H., et al. (1995) Structure of a new peptide, astin J, from *Aster tataricus*. *Chem. Pharm. Bull. (Tokyo)*, **43** (2), 271–273.
97. Itokawa, H., et al. (2000) Antitumor compounds isolated from higher plants. *J. Biochem. Mol. Biol. Biophys.*, **4** (3), 213–222.
98. Xu, H.M., et al. (2013) Astins K-P, six new chlorinated cyclopentapeptides from *Aster tataricus*. *Tetrahedron*, **69** (37), 7964–7969.
99. Xu, H.M., et al. (2013) Tataricins A and B, two novel cyclotetrapeptides from *Aster tataricus*, and their absolute configuration assignment. *Tetrahedron Lett.*, **54** (11), 1380–1383.
100. Jahn, L., et al. (2017) *Cyanodermella asteris* sp. nov. (Ostropales) from the inflorescence axis of *Aster tataricus*. *Mycotaxon Issn*, **132** (March), 107–123.
101. Jahn, L. (2015) Characterization of a new endophytic astin producer, *Pelliciarosea asterica*, from *Aster tataricus*. *PhD Thesis*, (October 2015), 1–154.
102. Schafhauser, T., et al. (2016) The cyclochlorotine mycotoxin is produced by the nonribosomal peptide synthetase CctN in *Talaromyces islandicus* ('*Penicillium islandicum*'). *Environ. Microbiol.*, **18** (11), 3728–3741.
103. Morita, H., et al. (1995) Thionation of an antitumour cyclic pentapeptide, astin B, from *Aster tataricus*. *Bioorganic Med. Chem. Lett.*, **5** (7), 677–680.
104. Morita, H., et al. (1996) Cyclic peptides from higher plants. XXVIII. Antitumor activity and hepatic microsomal biotransformation of cyclic pentapeptides, astins, from *Aster tataricus*. *Chem. Pharm. Bull. (Tokyo)*, **44** (5), 1026–1032.
105. Rossi, F., et al. (2004) New antitumour cyclic astin analogues: Synthesis, conformation and bioactivity. *J. Pept. Sci.*, **10** (2), 92–102.
106. Saviano, G., et al. (2004) Influence of conformational flexibility on biological activity in cyclic astin analogues. *Biopolym. - Pept. Sci. Sect.*, **76** (6), 477–484.
107. Cozzolino, R., et al. (2005) Antineoplastic cyclic astin analogues kill tumour cells via caspase-mediated induction of apoptosis. *Carcinogenesis*, **26** (4), 733–739.
108. Shen, Y., 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** (3), 260–268.

109. Grigoriev, I. V., et al. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.*, **42** (D1), D699–D704.
110. Leclère, V., et al. (2016) Bioinformatics Tools for the Discovery of New Nonribosomal Peptides, in *Methods in molecular biology (Clifton, N.J.)*, **1401**, 209–232.
111. Ochi, K. (2017) Insights into microbial cryptic gene activation and strain improvement: principle, application and technical aspects. *J. Antibiot. (Tokyo)*, **70** (1), 25–40.
112. Gardner, J.F., et al. (2009) Production of Citric Acid by Mutants of *Aspergillus niger*. *J. Gen. Microbiol.*, **14**, 228-237.
113. Newbert, R.W., et al. (1997) Analysis of a commercially improved *Penicillium chrysogenum* strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. *J. Ind. Microbiol. Biotechnol.*, **19** (1), 18–27.
114. Sung, C.T., et al. (2017) Overexpression of a three-gene conidial pigment biosynthetic pathway in *Aspergillus nidulans* reveals the first NRPS known to acetylate tryptophan. *Fungal Genet. Biol.*, **101**, 1–6.
115. Janevska, S., et al. (2017) Establishment of the inducible Tet-on system for the activation of the silent trichosetin gene cluster in *Fusarium fujikuroi*. *Toxins (Basel)*, **9** (4), 126.
116. Von Bargen, K.W., et al. (2013) Structure elucidation and antimalarial activity of apicidin F: An apicidin-like compound produced by *Fusarium fujikuroi*. *J. Nat. Prod.*, **76** (11), 2136-2140.
117. Palmer, J.M., and Keller, N.P. (2010) Secondary metabolism in fungi: Does chromosomal location matter? *Curr. Opin. Microbiol.*, **13** (4), 431–436.
118. Strauss, J., and Reyes-Dominguez, Y. (2011) Regulation of secondary metabolism by chromatin structure and epigenetic codes. *Fungal Genet. Biol.*, **48** (1), 62–69.
119. Soukup, A.A., et al. (2016) Enhancing nonribosomal peptide biosynthesis in filamentous fungi, in *Methods in Molecular Biology*, **1401**, 149–160.
120. Westphal, K.R., et al. (2019) Fusaoctaxin A, an Example of a Two-Step Mechanism for Non-Ribosomal Peptide Assembly and Maturation in Fungi. *Toxins (Basel)*, **11** (5), 277.
121. Perrin, R.M., et al. (2007) Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. *PLoS Pathog.*, **3** (4), e50.
122. Bok, J.W., and Keller, N.P. (2004) LaeA, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryot. Cell*, **3** (2), 527–535.
123. Bok, J.W., et al. (2006) Secondary metabolic gene cluster silencing in *Aspergillus nidulans*. *Mol. Microbiol.*, **61** (6), 1636–1645.
124. Zhang, A., et al. (2003) Efficient disruption of a polyketide synthase gene (pks1) required for melanin synthesis through *Agrobacterium*-mediated transformation of *Glarea lozoyensis*. *Mol. Genet. Genomics*, **268** (5), 645–655.
125. Chiang, Y.-M., et al. (2013) An efficient system for heterologous expression of secondary metabolite genes in *Aspergillus nidulans*. *J. Am. Chem. Soc.*, **135** (20), 7720–7731.
126. Tong, Z., et al. (2019) Systems metabolic engineering for citric acid production by *Aspergillus niger* in the post-genomic era. *Microb. Cell Fact.*, **18** (1), 1–15.
127. Souza, P.M., et al. (2017) Optimization and purification of l-asparaginase from fungi: A systematic review. *Crit. Rev. Oncol. Hematol.*, **120**, 194–202.
128. Masurekar, P.S., et al. (1992) Pneumocandins from *Zalerion arboricola*. II. Modification of product spectrum by mutation and medium manipulation. *J. Antibiot. (Tokyo)*, **45** (12), 1867–1874.
129. Ishidoh, K.I., et al. (2014) Identification of a gene cluster responsible for the biosynthesis of cyclic lipopeptide verlamelin. *Appl. Microbiol. Biotechnol.*, **98** (17), 7501-7510.
130. Biniarz, P., et al. (2018) High-throughput optimization of medium components and culture conditions for the efficient production of a lipopeptide pseudofactin by *Pseudomonas fluorescens* BD5. *Microb. Cell Fact.*, **17** (1).

131. Yaseen, Y., et al. (2017) Study of the correlation between fengycin promoter expression and its production by *Bacillus subtilis* under different culture conditions and the impact on surfactin production. *Arch. Microbiol.*, **199** (10), 1371–1382.
132. Baccile, J.A., et al. (2018) Fungal Isocyanide Synthases and Xanthocillin Biosynthesis in *Aspergillus fumigatus*. *MBio*, **9** (3), e00785-18.
133. Papagianni, M., and Matthey, M. (2004) Physiological aspects of free and immobilized *Aspergillus niger* cultures producing citric acid under various glucose concentrations. *Process Biochem.*, **39** (12), 1963–1970.
134. Villena, G.K., and Gutiérrez-Correa, M. (2007) Morphological patterns of *Aspergillus niger* biofilms and pellets related to lignocellulolytic enzyme productivities. *Lett. Appl. Microbiol.*, **45** (3), 231–237.
135. Mizunuma, T., et al. (2007) A mycelium with polyelectrolyte complex-bunched hyphae: preparation and fermentation performance. *Colloids Surf. B. Biointerfaces*, **56** (1–2), 155–160.
136. Pupin, M., et al. (2016) Norine: A powerful resource for novel nonribosomal peptide discovery. *Synth. Syst. Biotechnol.*, **1** (2), 89–94.
137. Hamada, T., et al. (2005) Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, *Theonella swinhoei*. *J. Am. Chem. Soc.*, **127** (1), 110–118.
138. Wilson, M.C., et al. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature*, **506** (7486), 58–62.
139. Freeman, M.F., et al. (2016) Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated “*Entotheonella*” bacteria. *Curr. Opin. Chem. Biol.*, **31**, 8–14.
140. Rausch, C., et al. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.*, **7**, 1–15.
141. Richter, L., et al. (2014) Engineering of *Aspergillus niger* for the production of secondary metabolites. *Fungal Biol. Biotechnol.*, **1** (1), 4.
142. Esmaeel, Q., et al. (2017) Nonribosomal peptides and polyketides of *Burkholderia*: new compounds potentially implicated in biocontrol and pharmaceuticals. *Environ. Sci. Pollut. Res.*, 1–14.
143. Kopp, F., and Marahiel, M.A. (2007) Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.*, **24** (4), 735.
144. Du, L., and Lou, L. (2010) PKS and NRPS release mechanisms. *Nat. Prod. Rep.*, **27** (2), 255–278.
145. Chiang, Y.M., et al. (2014) Analyzing Fungal Secondary Metabolite Genes and Gene Clusters. *Nat. Prod. Discourse, Divers. Des.*, 171–193.
146. Gibson, D.M., et al. (2014) Discovering the secondary metabolite potential encoded within entomopathogenic fungi. *Nat. Prod. Rep.*, **31** (10), 1287–1305.
147. Tang, M.C., et al. (2018) Identification of the pyranonigrin A biosynthetic gene cluster by genome mining in *Penicillium thymicola* IBT 5891. *AIChE J.*, **64** (12), 4182–4186.
148. Wang, X., et al. (2015) Genomic and transcriptomic analysis of the endophytic fungus reveals its lifestyle and high potential for synthesis of natural products. *BMC Genomics*, **16** (1), 1–13.
149. Kildgaard, S., et al. (2017) A dereplication and bioguided discovery approach to reveal new compounds from a marine-derived fungus *Stilbella fimetaria*. *Mar. Drugs.*, **15** (8).
150. Kunakom, S., and Eustáquio, A.S. (2019) Natural Products and Synthetic Biology: Where We Are and Where We Need To Go. *mSystems*, **4** (3), 1–4.
151. Luo, Y., et al. (2014) Recent advances in natural product discovery. *Curr. Opin. Biotechnol.*, **30**, 230–237.
152. Mushtaq, S., et al. (2018) Natural products as reservoirs of novel therapeutic agents. *EXCLI J.*, **17**, 420–451.

153. Anyaogu, D.C., and Mortensen, U.H. (2015) Heterologous production of fungal secondary metabolites in *Aspergilli*. *Front. Microbiol.*, **6** (FEB), 1–6.
154. Lazarus, C.M., et al. (2014) Reconstructing fungal natural product biosynthetic pathways. *Nat. Prod. Rep.*, **31** (10), 1339–1347.
155. Luo, Y., et al. (2016) New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.*, **33** (2), 174–82.
156. Weber, T., and Kim, H.U. (2016) The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.*, **1** (2), 69–79.
157. Schümann, J., and Hertweck, C. (2006) Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *J. Biotechnol.*, **124** (4), 690–703.
158. Zhang, H., et al. (2011) Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.*, **28** (1), 125–151.
159. Chooi, Y.-H., and Tang, Y. (2012) Navigating the fungal polyketide chemical space: from genes to molecules. *J. Org. Chem.*, **77** (22), 9933–9953.
160. Nah, H.-J., et al. (2017) Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces Species*. *Front. Microbiol.*, **8** (March), 394.
161. Tsunematsu, Y., et al. (2013) Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.*, **9** (12), 818–825.
162. Sung, C.T., et al. (2017) Overexpression of a three-gene conidial pigment biosynthetic pathway in *Aspergillus nidulans* reveals the first NRPS known to acetylate tryptophan. *Fungal Genet. Biol.*, **101**, 1–6.
163. Weber, T., et al. (2015) Metabolic engineering of antibiotic factories: New tools for antibiotic production in actinomycetes. *Trends Biotechnol.*, **33** (1), 15–26.
164. Beites, T., and Mendes, M. V. (2015) Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial secondary metabolism. *Front. Microbiol.*, **6** (SEP), 1–10.
165. Ziemert, N., et al. (2016) The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.*, **33** (8), 988–1005.
166. Yandell, M., and Ence, D. (2012) A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.*, **13** (5), 329–342.
167. Loman, N.J., et al. (2012) High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.*, **10** (9), 599–606.
168. Tatusova, T., et al. (2014) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.*, **42** (D1), D553–D559.
169. Heather, J.M., and Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics*, **107** (1), 1–8.
170. Koren, S., and Phillippy, A.M. (2015) One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.*, **23**, 110–120.
171. Vesth, T.C., et al. (2016) FunGeneClusterS: Predicting fungal gene clusters from genome and transcriptome data. *Synth. Syst. Biotechnol.*, **1** (2), 122–129.
172. Conway, K.R., and Boddy, C.N. (2013) ClusterMine360: A database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res.*, **41** (D1), 402–407.
173. Jain, M., et al. (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.*, **36** (4), 338–345.
174. Lu, H., et al. (2016) Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics. Proteomics Bioinformatics*, **14** (5), 265–279.

175. Rhoads, A., and Au, K.F. (2015) PacBio Sequencing and Its Applications. *Genomics. Proteomics Bioinformatics*, **13** (5), 278–289.
176. Kolmogorov, M., et al. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.*, **37**, 540-546.
177. Koren, S., et al. (2017) Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. *Genome Res.*, **27** (5), 722–736.
178. Bashir, A., et al. (2012) A hybrid approach for the automated finishing of bacterial genomes. *Nat. Biotechnol.*, **30** (7), 701-707.
179. Antipov, D., et al. (2016) HybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics*, **32** (7), 1009–1015.
180. Walker, B.J., et al. (2014) Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One*, **9** (11), e112963.
181. Chavali, A.K., and Rhee, S.Y. (2018) Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief. Bioinform.*, **19** (5), 1022–1034.
182. Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14** (9), 755–763.
183. Eddy, S.R. (2011) Accelerated Profile HMM Searches. *PLoS Comput. Biol.*, **7** (10), e1002195.
184. Blin, K., et al. (2017) Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief. Bioinform.*, 1-11.
185. Li, M.H.T., et al. (2009) Automated genome mining for natural products. *BMC Bioinformatics*, **10**, 185.
186. Starcevic, A., et al. (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.*, **36** (21), 6882–6892.
187. Khaldi, N., et al. (2010) SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.*, **47** (9), 736-741.
188. Blin, K., et al. (2017) AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.*, **45** (W1), W36–W41.
189. Wolf, T., et al. (2016) CASSIS and SMIPS: Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics.*, **32** (8), 1138-1143.
190. Challis, G.L., et al. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.*, **7** (3), 211–224.
191. Minowa, Y., et al. (2007) Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *J. Mol. Biol.*, **368** (5), 1500–1517.
192. Prieto, C., et al. (2012) NRPSp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics*, **28** (3), 426–427.
193. Khayatt, B.I., et al. (2013) Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models. *PLoS One*, **8** (4), e62136.
194. Rausch, C., et al. (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.*, **33** (18), 5799–5808.
195. Röttig, M., et al. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.*, **39** (suppl_2), 362–367.
196. Agüero-Chapin, G., et al. (2016) Alignment-Free Methods for the Detection and Specificity Prediction of Adenylation Domains, in *Methods in molecular biology (Clifton, N.J.)*, **1401**, 253–272.
197. Belshaw, P.J., et al. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284** (5413), 486-489.
198. Knudsen, M., et al. (2016) Computational discovery of specificity-conferring sites in non-

- ribosomal peptide synthetases. *Bioinformatics*, **32** (3), 325–329.
199. Chevrette, M.G., et al. (2017) SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics*, **33** (20), 3202–3210.
 200. Ziemert, N., et al. (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One*, **7** (3), e34064.
 201. Caradec, T., et al. (2014) Prediction of Monomer Isomery in Florine: A Workflow Dedicated to Nonribosomal Peptide Discovery. *PLoS One*, **9** (1), e85667.
 202. Farag, S., et al. (2019) Inter-Modular Linkers play a crucial role in governing the biosynthesis of non-ribosomal peptides. *Bioinformatics*, btz **127**.
 203. NCBI Resource Coordinators (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **44** (D1), 7-19.
 204. McGinnis, S., and Madden, T.L. (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.*, **32** (Web Server issue), 20-25.
 205. Blin, K., et al. (2017) The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.*, **45** (D1), D555–D559.
 206. Medema, M.H., et al. (2013) Detecting Sequence Homology at the Gene Cluster Level with MultiGeneBlast. *Mol. Biol. Evol.*, **30** (5), 1218–1223.
 207. Weber, T., et al. (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.*, **43** (W1), 237-243.
 208. Elliott, T.A., and Gregory, T.R. (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos. Trans. R. Soc. B Biol. Sci.*, **370** (1678).
 209. Setubal, J.C., et al. (2018) Comparative genomics for prokaryotes. *Methods in Molecular Biology*, **1704**, 55-78.
 210. Medema, M.H., et al. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*, **39** (Web Server issue), 339-346.
 211. Chutrakul, C., et al. (2008) The production and characterisation of trichotoxin peptaibols, by *Trichoderma asperellum*. *Chem. Biodivers.*, **5** (9), 1694–1706.
 212. Stajich, J.E. (2017) Fungal Genomes and Insights into the Evolution of the Kingdom, in *The Fungal Kingdom*, vol. 5, American Society of Microbiology, pp. 619–633.
 213. Araujo, R., and Sampaio-Maia, B. (2018) Fungal Genomes and Genotyping, in *Advances in applied microbiology*, vol. 102, pp. 37–81.
 214. Bode, H.B., et al. (2015) Structure Elucidation and Activity of KolossinA, the D -/ L - Pentadecapeptide Product of a Giant Nonribosomal Peptide Synthetase. *Angew. Chemie - Int. Ed.*, **54** (35), 10352–10355.
 215. Soukup, A.A., et al. (2016) Nonribosomal Peptide and Polyketide Biosynthesis. *Methods and protocols*, **1401**, 149–160.
 216. Skiba, M.A., et al. (2018) PKS–NRPS Enzymology and Structural Biology: Considerations in Protein Production. *Methods. Enzymol.*, **604**, 45-88.
 217. Tsunematsu, Y., et al. (2013) Yeast-based genome mining, production and mechanistic studies of the biosynthesis of fungal polyketide and peptide natural products. *Nat. Prod. Rep.*, **30** (8), 1139–1149.
 218. Mattanovich, D., et al. (2012) Recombinant Protein Production in Yeasts, in *Methods in molecular biology (Clifton, N.J.)*, vol. 824, pp. 329–358.
 219. Wehrs, M., et al. (2018) Production efficiency of the bacterial non-ribosomal peptide indigoidine relies on the respiratory metabolic state in *S. cerevisiae*. *Microb. Cell Fact.*, **17** (1), 193.
 220. Barajas, J.F., et al. (2017) Engineered polyketides: Synergy between protein and host level

- engineering. *Synth. Syst. Biotechnol.*, **2** (3), 147–166.
221. Saraya, R., et al. (2014) Tools for genetic engineering of the yeast *Hansenula polymorpha*. *Methods Mol. Biol.*, **1152**, 43–62.
222. Gidijala, L., et al. (2009) An engineered yeast efficiently secreting penicillin. *PLoS One*, **4** (12).
223. Kasahara, K., et al. (2010) Solanapyrone synthase, a possible Diels-Alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from *Alternaria solani*. *ChemBioChem*, **11** (9), 1245–1252.
224. Xue, Y., et al. (2017) Methylotrophic yeast *Pichia pastoris* as a chassis organism for polyketide synthesis via the full citrinin biosynthetic pathway. *J. Biotechnol.*, **242**, 64–72.
225. Erler, A., et al. (2006) Recombineering reagents for improved inducible expression and selection marker re-use in *Schizosaccharomyces pombe*. *Yeast*, **23** (11), 813–823.
226. Unkles, S.E., et al. (2014) Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem. Biol.*, **21** (4), 502–508.
227. Zhang, W., and Liu, J. (2016) Recent Advances in Understanding and Engineering Polyketide Synthesis. *F1000Research*, **5** (0), 208.
228. Ongley, S.E., et al. (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.*, **30** (8), 1121–1138.
229. Pickens, L., et al. (2014) Metabolic Engineering for the Production of Natural Products. *Annu. Rev. Chem. Biomol. Eng.*, **2** (3), 211–236.
230. Mattern, D.J., et al. (2015) Synthetic biology of fungal natural products. *Front. Microbiol.*, **6** (JUL), 1–7.
231. He, Y., et al. (2018) Recent advances in reconstructing microbial secondary metabolites biosynthesis in *Aspergillus* spp. *Biotechnol. Adv.*, **36** (3), 739–783.
232. Sakai, K., et al. (2008) Construction of a Citrinin Gene Cluster Expression System in Heterologous *Aspergillus oryzae*. *J. Biosci. Bioeng.*, **106** (5), 466–472.
233. Pahirulzaman, K.A.K., et al. (2012) A toolkit for heterologous expression of metabolic pathways in *aspergillus oryzae*. *Methods. Enzymol.*, **517**, 241–260
234. Bryksin, A., and Matsumura, I. (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques*, **48** (6), 463–465.
235. Horton, R.M., et al. (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **77** (1), 61–68.
236. Gibson, D.G., et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6** (5), 343–345.
237. Smith, C., et al. (1993) Generation of cohesive ends on PCR products by UDG-mediated excision of dU, and application for cloning into restriction digest-linearized vectors. *PCR Methods Appl.*, **2** (4), 328–332.
238. Cavaleiro, A.M., et al. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.*, **4** (9), 1042–1046.
239. Kouprina, N., and Larionov, V. (2016) Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma*, **125** (4), 621–632.
240. Zhang, J.J., et al. (2019) Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination, in *Methods in Enzymology*, vol. 621, Academic Press Inc., pp. 87–110.
241. Schimming, O., et al. (2014) Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *ChemBiochem*, **15** (9), 1290–1294.
242. Carter, M., and Shieh, J. (2015) Molecular Cloning and Recombinant DNA Technology, in *Guide to Research Techniques in Neuroscience*, Elsevier, pp. 219–237.
243. Collins, J., and Brüning, H.J. (1978) Plasmids useable as gene-cloning vectors in an in vitro packaging by coliphage λ : “cosmids.” *Gene*, **4** (2), 85–107.

244. Bok, J.W., et al. (2015) Fungal artificial chromosomes for mining of the fungal secondary metabolome. *BMC Genomics*, **16** (1), 1–10.
245. Clevenger, K.D., et al. (2017) A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat. Chem. Biol.*, **13** (8), 895–901.
246. Hughes, S.R., et al. (2015) Process for Assembly and Transformation into *Saccharomyces cerevisiae* of a Synthetic Yeast Artificial Chromosome Containing a Multigene Cassette to Express Enzymes That Enhance Xylose Utilization Designed for an Automated Platform. *J. Lab. Autom.*, **20** (6), 621–635.
247. Huo, L., et al. (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.*, **19** (10), 1278–1287.
248. Kosuri, S., and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods*, **11** (5), 499–507.
249. Li, C., et al. (2017) Recycling of a selectable marker with a self-excisable plasmid in *Pichia pastoris*. *Sci. Rep.*, **7** (1), 11113.
250. Kumakura, N., et al. (2019) Establishment of a selection marker recycling system for sequential transformation of the plant-pathogenic fungus *Colletotrichum orbiculare*. *Mol. Plant Pathol.*, **20** (3), 447–459.
251. Wasil, Z., et al. (2013) One pathway, many compounds: Heterologous expression of a fungal biosynthetic pathway reveals its intrinsic potential for diversity. *Chem. Sci.*, **4** (10), 3845–3856.
252. Engler, C., et al. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, **3** (11).
253. Terfrüchte, M., et al. (2014) Establishing a versatile Golden Gate cloning system for genetic engineering in fungi. *Fungal Genet. Biol.*, **62**, 1–10.
254. Agmon, N., et al. (2015) Yeast Golden Gate (yGG) for the Efficient Assembly of *S. cerevisiae* Transcription Units. *ACS Synth. Biol.*, **4** (7), 853–859.
255. Mitchell, L.A., et al. (2015) Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res.*, **43** (13), 6620–6630.
256. Kakui, Y., et al. (2015) Module-based construction of plasmids for chromosomal integration of the fission yeast *Schizosaccharomyces pombe*. *Open Biol.*, **5** (6), 150054.
257. Celińska, E., et al. (2017) Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. *Microb. Biotechnol.*, **10** (2), 450–455.
258. Kim, J.H., et al. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*, **6** (4), e18556.
259. Boeke, J.D., et al. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.*, **154**, 164–175.
260. Akada, R., et al. (2006) PCR-mediated seamless gene deletion and marker recycling in *Saccharomyces cerevisiae*. *Yeast*, **23** (5), 399–405.
261. Sauer, B. (1987) Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **7** (6), 2087–2096.
262. Pan, R., et al. (2011) Sequential deletion of *Pichia pastoris* genes by a self-excisable cassette. *FEMS Yeast Res.*, **11** (3), 292–298.
263. Yaegashi, J., et al. (2014) Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in *Aspergillus nidulans*. *J Ind Microbiol Biotechnol.*, **41** (2), 433–442.
264. van Dijk, J.W.A., and Wang, C.C.C. (2016) Heterologous Expression of Fungal Secondary Metabolite Pathways in the *Aspergillus nidulans* Host System, in *Methods in Enzymology*.
265. Gidijala, L., et al. (2008) Production of functionally active *Penicillium chrysogenum* isopenicillin N synthase in the yeast *Hansenula polymorpha*. *BMC Biotechnol.*, **8**, 29.
266. Frisvad, J.C., et al. (2009) Metabolomics of *Aspergillus fumigatus*. *Med. Mycol.*, **47** (SUPPL).

- 1), 53–71.
267. Ames, B.D., et al. (2010) Enzymatic processing of fumiquinazoline F: A tandem oxidative-acylation strategy for the generation of multicyclic scaffolds in fungal indole alkaloid biosynthesis. *Biochemistry*, **49** (39), 8564–8576.
268. Ishiuchi, K., et al. (2012) Establishing a New Methodology for Genome Mining and Biosynthesis of Polyketides and Peptides through Yeast Molecular Genetics. *ChemBioChem*, **13** (6), 846–854.
269. Haynes, S.W., et al. (2013) Complexity generation in fungal peptidyl alkaloid biosynthesis: A two-enzyme pathway to the hexacyclic MDR export pump inhibitor ardeemin. *ACS Chem. Biol.*, **8** (4), 741–748.
270. Gao, X., et al. (2013) An iterative, bimodular nonribosomal peptide synthetase that converts anthranilate and tryptophan into tetracyclic asperlicins. *Chem. Biol.*, **20** (7), 870–878.
271. Yu, D., et al. (2017) Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.*, **8** (May), 1–11.
272. Yu, D., et al. (2013) Engineered production of fungal anticancer cyclooligomer depsipeptides in *Saccharomyces cerevisiae*. *Metab. Eng.*, **18**, 60–68.
273. Xu, W., et al. (2010) Analysis of intact and dissected fungal polyketide synthase-nonribosomal peptide synthetase in vitro and in *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.*, **132** (39), 13604–13607.
274. Hühner, E., et al. (2018) Production of α -keto carboxylic acid dimers in yeast by overexpression of NRPS-like genes from *Aspergillus terreus*. *Appl. Microbiol. Biotechnol.*, **102** (4), 1663–1672.
275. Wang, M., and Zhao, H. (2014) Characterization and engineering of the adenylation domain of a NRPS-like protein: A potential biocatalyst for aldehyde generation. *ACS Catal.*, **4** (4), 1219–1225.
276. Bozhüyük, K.A.J., et al. (2019) Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.*, **11** (7), 653–661.
277. Siewers, V., et al. (2010) Implementation of communication-mediating domains for non-ribosomal peptide production in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.*, **106** (5), 841–844.
278. Smith, D.J., et al. (1990) Beta-lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. *EMBO J.*, **9** (3), 741–747.
279. Felenbok, B., et al. (2004) Ethanol catabolism in *Aspergillus nidulans*: A model system for studying gene regulation. **69**, 149–204.
280. Wunsch, C., et al. (2015) Targeted production of secondary metabolites by coexpression of non-ribosomal peptide synthetase and prenyltransferase genes in *Aspergillus*. *Appl. Microbiol. Biotechnol.*, **99** (10), 4213–4223.
281. Munawar, A., et al. (2013) Isolation and Characterisation of a Ferrirhodin Synthetase Gene from the Sugarcane Pathogen *Fusarium sacchari*. *ChemBioChem*, **14** (3), 388–394.
282. Brandenburger, E., et al. (2017) A Highly Conserved Basidiomycete Peptide Synthetase Produces a Trimeric Hydroxamate Siderophore. *Appl. Environ. Microbiol.*, **83** (21), 1–13.
283. Li, W., et al. (2018) Asperphenamate biosynthesis reveals a novel two-module NRPS system to synthesize amino acid esters in fungi. *Chem. Sci.*, **9** (9), 2589–2594.
284. Chang, S.L., et al. (2013) Reconstitution of the early steps of gliotoxin biosynthesis in *Aspergillus nidulans* reveals the role of the monooxygenase GliC. *Bioorganic Med. Chem. Lett.*, **23** (7), 2155–2157.
285. De Mattos-Shipley, K.M.J., et al. (2018) The cycloaspeptides: Uncovering a new model for methylated nonribosomal peptide biosynthesis. *Chem. Sci.*, **9** (17), 4109–4117.
286. Yoshimi, A., et al. (2018) Heterologous production of a novel cyclic peptide compound, KK-1, in *Aspergillus oryzae*. *Front. Microbiol.*, **9** (APR), 1–12.

287. Guo, C.J., et al. (2015) Spatial regulation of a common precursor from two distinct genes generates metabolite diversity. *Chem. Sci.*, **6** (10), 5913–5921.
288. Geib, E., et al. (2016) A Non-canonical Melanin Biosynthesis Pathway Protects *Aspergillus terreus* Conidia from Environmental Stress. *Cell Chem. Biol.*, **23** (5), 587–597.
289. Geib, E., and Brock, M. (2017) ATNT: an enhanced system for expression of polycistronic secondary metabolite gene clusters in *Aspergillus niger*. *Fungal Biol. Biotechnol.*, **4** (1), 13.
290. Yeh, H.-H., et al. (2012) Molecular genetic analysis reveals that a nonribosomal peptide synthetase-like (NRPS-like) gene in *Aspergillus nidulans* is responsible for microperfurane biosynthesis. *Appl. Microbiol. Biotechnol.*, **96** (3), 739–748.
291. Sivanathan, S., and Scherckenbeck, J. (2014) Cyclodepsipeptides: A Rich Source of Biologically Active Compounds for Drug Research. *Molecules*, **19** (8), 12368–12420.
292. Boecker, S., et al. (2018) *Aspergillus niger* is a superior expression host for the production of bioactive fungal cyclodepsipeptides. *Fungal Biol. Biotechnol.*, **5** (1), 4.
293. Meyer, V., et al. (2011) Fungal Gene Expression on Demand: an Inducible, Tunable, and Metabolism-Independent Expression System for *Aspergillus niger*. *Appl. Environ. Microbiol.*, **77** (9), 2975–2983.
294. Brown, A.S., et al. (2018) Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.*
295. Zobel, S., et al. (2016) Reprogramming the Biosynthesis of Cyclodepsipeptide Synthetases to Obtain New Enniatins and Beauvericins. *ChemBioChem*, **17** (4), 283–287.
296. Steiniger, C., et al. (2017) Harnessing fungal nonribosomal cyclodepsipeptide synthetases for mechanistic insights and tailored engineering. *Chem. Sci.*, **8** (11), 7834–7843.
297. Süßmuth, R.D., and Mainz, A. (2017) Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew. Chemie - Int. Ed.*, **56** (14), 3770–3821.
298. Kakule, T.B., et al. (2014) Combinatorialization of fungal polyketide synthase-peptide synthetase hybrid proteins. *J. Am. Chem. Soc.*, **136** (51), 17882–17890.
299. Nielsen, M.L., et al. (2016) Linker flexibility facilitates module exchange in fungal hybrid PKS-NRPS engineering. *PLoS One*, **11** (8), 1–18.
300. Fisch, K.M., et al. (2011) Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.*, **133** (41), 16635–16641.
301. Halo, L.M., et al. (2008) Authentic heterologous expression of the tenellin iterative polyketide synthase nonribosomal peptide synthetase requires coexpression with an enoyl reductase. *ChemBioChem*, **9** (4), 585–594.
302. Heneghan, M.N., et al. (2010) First heterologous reconstruction of a complete functional fungal biosynthetic multigene cluster. *ChemBioChem*, **11** (11), 1508–1512.
303. Sato, M., et al. (2015) Involvement of Lipocalin-like CghA in Decalin-Forming Stereoselective Intramolecular [4+2] Cycloaddition. *ChemBioChem*, **16** (16), 2294–2298.
304. Song, Z., et al. (2015) Heterologous expression of the avirulence gene ACE1 from the fungal rice pathogen *Magnaporthe oryzae*. *Chem. Sci.*, **6** (8), 4837–4845.
305. Qun Yue, et al. (2015) Functional Operons in Secondary Metabolic Gene Clusters in *Glarea lozoyensis* (Fungi, Ascomycota, Leotiomycetes). *MBio*, **6** (3).
306. Tanaka, M., et al. (2014) Effects of codon optimization on the mRNA levels of heterologous genes in filamentous fungi. *Appl. Microbiol. Biotechnol.*, **98** (9), 3859–3867.

307. Samel, S.A., et al. (2007) Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure*, **15** (7), 781–792.
308. Tarry, M.J., et al. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*, **25** (5), 783–793.e4.
309. Kealey, J.T., et al. (1998) Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc. Natl. Acad. Sci. U. S. A.*, **95** (2), 505–509.
310. Jones, E.W. (1991) Tackling the protease problem in *Saccharomyces cerevisiae*. *Methods Enzymol.*, **194**, 428–453.
311. Siewers, V., et al. (2009) Heterologous production of non-ribosomal peptide LLD-ACV in *Saccharomyces cerevisiae*. *Metab. Eng.*, **11** (6), 391–397.
312. Mootz, H.D., et al. (2002) Functional characterization of 4'-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of *Saccharomyces cerevisiae* lys5. *FEMS Microbiol. Lett.*, **213** (1), 51–57.
313. Mofid, M.R., et al. (2002) Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp. *J. Biol. Chem.*, **277** (19), 17023–17031.
314. Tobias, N.J., et al. (2016) *Legionella* shows a diverse secondary metabolism dependent on a broad spectrum Sfp-type phosphopantetheinyl transferase. *PeerJ*, **4**, e2720.
315. Awan, A.R., et al. (2017) Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nat. Commun.*, **8** (May), 1–8.
316. Gardiner, D.M., and Howlett, B.J. (2005) Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*. *FEMS Microbiol. Lett.*, **248** (2), 241–248.
317. Vassaux, A., et al. (2019) Astin C production by the endophytic fungus *Cyanodermella asteris* in planktonic and immobilized culture conditions. *Biotechnol. J.*, **14** (8), e1800624.
318. Inglis, D.O., et al. (2013) Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae*. *BMC Microbiol.*, **13** (1).
319. Bond, C., et al. (2016) *Saccharomyces cerevisiae* as a tool for mining, studying and engineering fungal polyketide synthases. *Fungal Genet. Biol.*, **89**, 52–61.
320. Martín, J.F., et al. (2005) Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. *Curr. Opin. Microbiol.*, **8** (3), 282–293.
321. Chanda, A., et al. (2009) A key role for vesicles in fungal secondary metabolism. *Proc. Natl. Acad. Sci.*, **106** (46), 19533–19538.
322. Nielsen, J.C., and Nielsen, J. (2017) Development of fungal cell factories for the production of secondary metabolites: Linking genomics and metabolism. *Synth. Syst. Biotechnol.*, **2** (1), 5–12.
323. Fujii, I., et al. (1996) Cloning of the polyketide synthase gene atX from *Aspergillus terreus* and its identification as the 6-methylsalicylic acid synthase gene by heterologous expression. *Mol. Gen. Genet.*, **253** (1–2), 1–10.
324. Cundliffe, E., and Demain, A.L. (2010) Avoidance of suicide in antibiotic-producing microbes. *J. Ind. Microbiol. Biotechnol.*, **37** (7), 643–672.
325. Keller, N.P. (2015) Translating biosynthetic gene clusters into fungal armor and weaponry. *Nat. Chem. Biol.*, **11** (9), 671–677.
326. do Valle Matta, M.A., et al. (2001) Novel target genes of the yeast regulator Pdr1p: a contribution of the TPO1 gene in resistance to quinidine and other drugs. *Gene*, **272** (1–2), 111–9.
327. Dolan, S.K., et al. (2015) Resistance is not futile: gliotoxin biosynthesis, functionality and utility. *Trends Microbiol.*, **23** (7), 419–428.
328. Sharma, C., and Chowdhary, A. (2017) Molecular bases of antifungal resistance in filamentous

- fungi. *Int. J. Antimicrob. Agents*, **50** (5), 607–616.
329. Tran, P.N., et al. (2019) Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl. Microbiol. Biotechnol.*, **103** (8), 3277–3287.
330. Rodriguez, S., et al. (2014) Production and quantification of sesquiterpenes in *Saccharomyces cerevisiae*, including extraction, detection and quantification of terpene products and key related metabolites. *Nat. Protoc.*, **9** (8), 1980–1996.
331. Singh, M.P., et al. (2010) Fermentative production of self-toxic fungal secondary metabolites. *J. Ind. Microbiol. Biotechnol.*, **37** (4), 335–340.
332. Fickers, P., et al. (2008) Temperature dependence of mycosubtilin homologue production in *Bacillus subtilis* ATCC6633. *Res. Microbiol.*, **159** (6), 449–457.
333. Bumpus, S.B., et al. (2009) A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.*, **27** (10), 951–956.
334. Schwarzer, D., et al. (2001) Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chem. Biol.*, **8** (10), 997–1010.
335. Phelan, V. V., et al. (2009) Adenylation Enzyme Characterization Using γ -18O4-ATP Pyrophosphate Exchange. *Chem. Biol.*, **16** (5), 473–478.
336. Duckworth, B.P., et al. (2016) Measurement of nonribosomal peptide synthetase adenylation domain activity using a continuous hydroxylamine release assay, in *Methods in Molecular Biology*, vol. 1401, Humana Press Inc., pp. 53–61.
337. Suárez, T., and Peñalva, M.A. (1996) Characterization of a *Penicillium chrysogenum* gene encoding a PacC transcription factor and its binding sites in the divergent pcbAB-pcbC promoter of the penicillin biosynthetic cluster. *Mol. Microbiol.*, **20** (3), 529–540.
338. Ball, A.P., et al. (1978) The Natural Penicillins — Benzylpenicillin (Penicillin G) and Phenoxymethylpenicillin (Penicillin V), in *Antibacterial Drugs Today*, Springer Netherlands, pp. 6–18.
339. Stöckmann, C., et al. (2009) Process development in *Hansenula polymorpha* and *Arxula adenivorans*, a re-assessment. *Microb. Cell Fact.*, **8** (1), 22.
340. Spröte, P., et al. (2009) Contribution of Peroxisomes to Penicillin Biosynthesis in *Aspergillus nidulans*. *Eukaryot. Cell*, **8** (3), 421–423.
341. Maury, J., et al. (2016) EasyCloneMulti: A set of vectors for simultaneous and multiple genomic integrations in *Saccharomyces cerevisiae*. *PLoS One*, **11** (3), 1–22.
342. Newbert, R.W., et al. (1997) Analysis of a commercially improved *Penicillium chrysogenum* strain series: Involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. *J. Ind. Microbiol. Biotechnol.*, **19** (1), 18–27.
343. Vournakis, J.N., and Elander, R.P. (1983) Genetic manipulation of antibiotic-producing microorganisms. *Science*, **219** (4585), 703–709.
344. Fierro, F., et al. (2006) Transcriptional and bioinformatic analysis of the 56.8 kb DNA region amplified in tandem repeats containing the penicillin gene cluster in *Penicillium chrysogenum*. *Fungal Genet. Biol.*, **43** (9), 618–629.
345. Müller, W.H., et al. (2018) Localization of the pathway of the penicillin biosynthesis in *Penicillium chrysogenum*. *EMBO J.*, **10** (2), 489–495.
346. Salo, O. V., et al. (2015) Genomic mutational analysis of the impact of the classical strain improvement program on β -lactam producing *Penicillium chrysogenum*. *BMC Genomics*, **16** (1).
347. Salo, O., et al. (2016) Identification of a polyketide synthase involved in sorbicillin biosynthesis by *Penicillium chrysogenum*. *Appl. Environ. Microbiol.*, **82** (13), 3971–3978.
348. Harris, D.M., et al. (2009) Exploring and dissecting genome-wide gene expression responses of *Penicillium chrysogenum* to phenylacetic acid consumption and penicillinG production. *BMC Genomics*, **10**.
349. Harvey, C.J.B., et al. (2018) HEx: A heterologous expression platform for the discovery of

fungal natural products. *Sci. Adv.*, **4** (4), eaar5459.

Chapitre 2

**Optimisation de la production homologue
d'astine C par le champignon endophyte
Cyanoderrella asteris (Article)**

- **Mise en contexte**

Jusqu'à présent, l'astine C était obtenue par extraction à partir des racines de la plante *Aster tataricus*. Ce procédé nécessite une très grande quantité de matériel végétal et de nombreuses étapes d'extraction / purification pour un rendement en astine C très faible. L'isolement récent de *Cyanodermella asteris*, un champignon endophyte de cette plante produisant de l'astine C, ouvre de nouvelles perspectives en matière de production. Néanmoins, son taux de croissance très faible dans les milieux couramment utilisés pour la culture de champignons filamenteux (extrait de malt, extrait de pomme de terre, etc.) limite sa productivité. De plus, à l'heure actuelle, il n'existe aucun outil moléculaire permettant d'accroître sa productivité par ingénierie génétique. Par conséquent, l'approche la plus évidente consiste à optimiser les conditions de culture de ce champignon afin de favoriser sa production. Dans ce contexte, le développement d'un système de culture où les cellules sont immobilisées sur un support, soit un état plus proche des conditions naturelles de développement de *C. asteris* (sur les racines d'*A. tataricus*), est une des pistes considérées.

- **Résumé**

Cyanodermella asteris a récemment été isolé à partir de la plante médicinale *Aster tataricus*. Ce champignon filamenteux produit de l'astine C, un pentapeptide cyclique aux propriétés anticancéreuses et anti-inflammatoires. Les taux de production de ce métabolite secondaire ont été comparés en cultivant le champignon en condition planctonique et sous forme de cellules immobilisées sur support. En condition de cellules immobilisées, un support en acier inoxydable était immergé dans le milieu de culture. Un développement fongique a été observé uniquement sur le support ce qui constitue un avantage considérable pour l'extraction et la purification de l'astine C sécrétée dans le milieu de culture. Plusieurs paramètres relatifs au métabolisme de *C. asteris*, tels que le taux de consommation du substrat, la croissance cellulaire, et la production d'astine C, ont été affectés par les conditions de culture. En cultivant le champignon sur un support, une augmentation de 30 % de la production d'astine C, ainsi qu'une augmentation de 39 % de la biomasse, ont été observées. Plusieurs paramètres de culture ont ensuite été considérés pour optimiser le système de production sur support. En particulier, le mode d'inoculation à partir de spores plutôt que d'hyphes, ainsi qu'un pré-lavage du support à l'hydroxyde de sodium, sont deux paramètres qui se sont avérés favorables à la fois au développement de *C. asteris* sur le support, et à la production d'astine C. L'influence de la composition et du pH du milieu de culture sur la production d'astine C a également été évaluée. Les conditions de culture optimisées ont permis d'obtenir un rendement spécifique en astine C cinq fois supérieur à celui précédemment décrit pour l'extraction d'astine C à partir d'*Aster tataricus* (0,89 et 0,16 mg/g respectivement).

- **Mots Clés**

Astine C, Biofilm, *Cyanodermella asteris*, Culture en cellules immobilisées, Métabolites secondaires

- **Principales conclusions**

Dans un premier temps, les taux de croissance et de production en astine C de *Cyanodermella asteris* ont été évalués en milieux à l'extrait de malt (MEA) dont la composition a été altérée (en modifiant la source de sucres et en supplémentant le milieu avec certains acides aminés). A partir de ces essais préliminaires, un milieu (MEAb) a été formulé dans le but de favoriser à la fois la croissance du champignon et la production en astine C. Ce milieu varie du milieu MEA par l'ajout de β -phénylalanine, et l'utilisation du mannose en tant que source de sucre.

Ce milieu a été utilisé pour cultiver *Cyanodermella asteris* à la fois en condition planctonique et sur un support. L'utilisation d'un système de culture où les cellules sont immobilisées a permis d'augmenter de 39 % le développement de la biomasse et de 30 % la production en astine C. Dans ces conditions, le champignon se développe exclusivement sur le support et pas dans le milieu liquide ce qui favorise la procédure d'extraction / purification de l'astine C qui est sécrétée dans le milieu de culture.

Afin d'optimiser de nouveau ce système, l'influence de plusieurs paramètres de culture sur la production d'astine C et le développement de *C. asteris* ont été évalués. En particulier, un pré-traitement du support avec de l'hydroxyde de sodium s'est avéré nécessaire pour le maintien de la productivité en astine C. Concernant le mode d'inoculation, l'utilisation de spores plutôt que d'hyphes, semble favoriser l'attachement et le développement de *C. asteris* sur le support. De plus, l'impact du pH du milieu de culture a également été évalué. La croissance du champignon et la production d'astine C ont été négativement impactées par le maintien du pH à 4,8. Une acidification progressive du milieu semble donc favoriser la production du métabolite. Enfin, *C. asteris* a également été cultivé sur support, en conditions optimisées, dans un milieu alternatif enrichi en sources de carbone. En combinant les différents paramètres considérés lors de l'optimisation du système, la production en astine C a pu être augmentée de 120 % par rapport aux conditions de culture initiales avec les cellules immobilisées.

Cet accroissement significatif constitue un premier pas dans le processus d'optimisation de la production d'astine C. Néanmoins, les rendements obtenus sont encore loin d'atteindre ceux attendus pour une production à échelle industrielle de ce composé d'intérêt. Les conditions de culture développées lors de cette étude peuvent constituer une base sur laquelle travailler pour optimiser de nouveau la production d'astine C à partir de *C. asteris*. Étant donné les faibles taux de croissance et de production observés, le transfert de la voie de biosynthèse de l'astine au sein d'un hôte hétérologue, aux conditions de culture plus adaptées aux méthodes de fermentation industrielles, est une approche envisageable.

Biotechnology Journal, 2019

Astin C production by the endophytic fungus *Cyanodermella asteris* in planktonic and immobilized culture conditions

Antoine Vassaux^{1,2}, Cédric Tarayre¹, Anthony Arguëlles-Arias¹, Philippe Compère³, Frank Delvigne¹, Patrick Fickers¹, Linda Jahn⁴, Alexander Lang⁵, Valérie Leclère², Jutta Ludwig-Müller⁴, Marc Ongena¹, Thomas Schafhauser⁶, Samuel Telek¹, Ariane Théatre¹, Willem J.H. van Berkel⁷, Micheline Vandenbol⁸, Karl-Heinz van Pée⁵, Luc Willems^{1,9}, Wolfgang Wohlleben⁶, Philippe Jacques¹

¹TERRA Teaching and Research Centre, Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d'Agronomie, B-5030, Gembloux, Belgium

²Univ. Lille, INRA, ISA, Univ. Artois, Univ. Littoral Côte d'Opale, EA 7394-ICV-Institut Charles Viollette, F-59000 Lille, France

³Département de Biologie, Ecologie et Evolution et Cellule d'Appui à la Recherche et à l'Enseignement en Microscopie, Université de Liège, Allée du Six Août 15, B-4000 Liège, Belgium

⁴Institut für Botanik, Technische Universität Dresden, 01062 Dresden, Germany

⁵Allgemeine Biochemie, Technische Universität Dresden, 01069 Dresden, Germany

⁶Mikrobiologie/Biotechnologie, Interfakultäres Institut für Mikrobiologie und Infektionsmedizin, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany

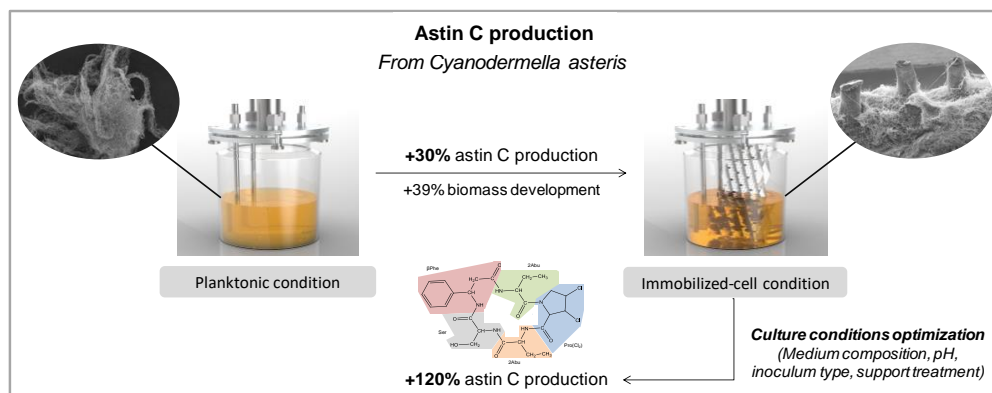
⁷Laboratory of Biochemistry, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

⁸TERRA Teaching and Research Centre, Microbiologie et Génomique, Gembloux Agro-Bio Tech, University of Liege, Avenue de la Faculté d'Agronomie, B-5030, Gembloux, Belgium

⁹GIGA Research, Cellular and Molecular epigenetics, University of Liege, allée de l'hôpital 1, B-4000 Liège, Belgium

Graphical abstract

The production of astin C, a promising anticancer and anti-inflammatory compound initially isolated from *Aster tataricus*, has been recently attributed to its fungal endophyte *Cyanoderrella asteris*. This study is the first to be dedicated to the optimization of the production of astin C by this fungus. Growing this fungus under immobilized-cell condition increases astin C production and facilitates its purification.



Abstract

The fungal endophyte *Cyanoderrella asteris* (*C. asteris*) has been recently isolated from the medicinal plant *Aster tataricus* (*A. tataricus*). This fungus produces astin C, a cyclic pentapeptide with anticancer and anti-inflammatory properties. The production of this secondary metabolite is compared in immobilized and planktonic conditions. For immobilized cultures, a stainless steel packing immersed in the culture broth is used as a support. In these conditions, the fungus exclusively grows on the packing, which provides a considerable advantage for astin C recovery and purification. *C. asteris* metabolism is different according to the culture conditions in terms of substrate consumption rate, cell growth, and astin C production. Immobilized-cell cultures yield a 30% increase of astin C production, associated with a 39% increase in biomass. The inoculum type as spores rather than hyphae, and a pre-inoculation washing procedure with sodium hydroxide, turns out to be beneficial both for astin C production and fungus development onto the support. Finally, the influence of culture parameters such as pH and medium composition on astin C production is evaluated. With optimized culture conditions, astin C yield is further improved reaching a five times higher final specific yield compared to the value reported with astin C extraction from *A. tataricus* (0.89 mg/g and 0.16 mg/g respectively).

Keywords: Astin C, Biofilm, *Cyanoderrella asteris*, Immobilized-cell culture, Secondary metabolite

1. Introduction

Filamentous ascomycetes fungi are known to have a highly developed secondary metabolism involved in the synthesis of a gold mine of natural compounds [1,2]. Recently, a newly described fungal endophyte has been isolated from the inflorescences of *Aster tataricus* (*A. tataricus*) [3], a fundamental herb in traditional Chinese and Japanese medicine, well-known for its antibiotic, antiviral, diuretic and anti-ulcer activities [4,5]. This endophyte, named *Cyanodermella asteris* (*C. asteris*), is related to the Ostropales fungi order belonging to the ascomycota phylum. Its genome analysis revealed the existence of 41 putative gene clusters involved in secondary metabolism [6]. Among them, 19 putative gene clusters related to the biosynthesis of Non-Ribosomal Peptide Synthetases (NRPSs) and Polyketide Synthases were identified.

One of the NRPS clusters was recently identified being involved in the biosynthesis of astin C, a di-chlorinated cyclopeptide originally detected in *A. tataricus* tissues [7]. This work provided unequivocal evidence for the production of astin C by the endophyte rather than by the plant as initially described [8,9]. Structurally, astin C is composed of two L-2-aminobutyrate, one L- β -phenylalanine (β Phe), one L-cis-3,4-dichloroproline, and one L-serine. Astin C has previously shown very promising activities such as antineoplastic, antitumor and anti-inflammatory effects [5,9,10]. Up to now, astin C was extracted and purified from *A. tataricus* roots. This operation requires huge amount of plant material to obtain very low yield of purified astin C (8 g from 50 kg of roots, i.e. 0.16 mg/g) [11]. Very recently, a multiple-steps approach to chemically synthesized astin C was reported [12]. The high costs of precursors, especially related to the unusual dichloroproline, prevent to reach substantial industrial astin C production yields. The identification of *C. asteris* as the real astin producer opens a new field of investigation to produce this compound in large scale by bioreactor process.

Given the ability of fungi to develop onto surfaces, it is likely that *C. asteris* growth could be achieved on a support in immobilized-cell cultures (ICC) system. Adherence of fungal cells on metallic structured packing in ICCs have been reported in flask and bioreactor systems [13]. ICCs have already been shown beneficial for the production rate of metabolites and enzymes by filamentous fungi [14]. Especially for *Aspergillus niger*, improvements were observed for the production of citric acid [15], gluconic acid [16], a set of glycosyl hydrolases [17] and a heterologous laccase from *Pleurotus ostreatus* fused to the catalytic part of glucoamylase [18]. Moreover, this production process could facilitate metabolites recovery by promoting the secretion of secondary metabolites in culture media [14].

In this research, upon optimization of the culture medium, we investigated the cell growth and astin C production by *C. asteris* in both planktonic and cell-immobilized conditions. This work first emphasizes on the feasibility and the benefits of growing *C. asteris* in ICCs. As the ICC was successful for astin C production, different support preparations were applied to the stainless steel support to identify the most favorable preparation technique to obtain high growth rate and high astin C yield. Finally, some

cultivation parameters such as inoculum type, medium composition and cultivation pH were considered in order to further optimize the astin C production

2. Experimental section

2.1. Fungal strain

C. asteris (Mycobank[®] accession number #814158), belonging to the order Ostropales (class Lecanoromycetes), was isolated from the plant *A. tataricus*, according to the technique described by Jahn et al. [3].

2.2. General cultivation conditions

The malt extract agar (MEA) growth media (composition: malt extract 30 g/L, glucose 20 g/L, meat peptone 1 g/L, trace elements), commonly used to cultivate filamentous fungi [19,20], was used for *C. asteris* culture maintenance. Each cultivation condition of *C. asteris* was performed in triplicates at 22 or 25 °C in flasks under shaking at 150-160 rpm. Initial media optimization experiments were performed in 50 mL culture volume inoculated with 0.25 mL of homogenized 7-day-old pre-culture. Per experiment, one medium ingredient was altered before autoclaving (amount and type of sugar) or additionally added after autoclaving (amino acids, sterile-filtered). Final concentrations of added amino acids were 0.5 g/L when appearing once in the final product astin C (L-serine, L-proline, L- β -phenylalanine) or 1 g/L when appearing twice (L-2-aminobutyric acid). Comparative experiments between planktonic and immobilized-cell conditions were performed in 100 mL of MEAb culture medium or 5 times diluted MEAb (d5MEAb) (composition: malt extract 10 g/L, mannose 20 g/L, meat peptone 1 g/L, β -phenylalanine 0.5 g/L, trace elements) inoculated with 1 mL of a three week-old pre-culture. A further medium, MEAc (composition: malt extract 20 g/L, glucose 20 g/L, mannose 30 g/L, meat peptone 1 g/L, trace elements), was used in the section related to astin C production optimization in immobilized-cell condition. All chemicals were provided by Sigma-Aldrich (St. Louis, MO, USA).

2.3. Statistical methods

The significance of the presented results was assessed through a one or two-way variance analysis using MINITAB. The application conditions were checked (normality test of Ryan-Joiner and test for equal variances) before making the variance analysis. Results were considered to be significantly different as soon as the p-value was below 0.05.

2.4. Immobilized-cell cultures

In immobilized-cell cultures a solid support, named packing, was partially immersed in the culture broth. Packings were constituted of two layers of stainless steel waved trellis (4.5 x 8.5 cm) (Sulzer Chemtech AG, Switzerland) joint together with plastic rings as described by Zune et al. [14]. Prior utilization, supports were immersed in 20% NaOH (24 h), before being rinsed with distilled water until all NaOH was removed (estimated by the neutral pH of the last washing water). Supports and culture media were sterilized separately. During 28 culture-days pH, insoluble dry

matter, astin C concentration and reducing sugar consumption were monitored. Reducing sugar quantification was performed via the dinitrosalicylic acid (DNS) technique [21].

2.5. Effect of the inoculum type on immobilized-cell cultures of *C. asteris*

Two types of *C. asteris* inocula were tested. The first inoculum (mycelium) consisted of 1 mL of hyphae from a 15 day-old culture in the MEAb medium. The second inoculum (spores) consisted of 1 mL of a mixture of hyphae and spores (1.16×10^6 spores/mL) obtained from a 55 day-old culture in MEAb. Microscopic observations of this culture revealed enrichment in spores compared to the 15 day-old culture. Cultures were maintained for 49 days. Cells attached on the solid support were removed by scratching off in order to determine dry matters. In some cultures, *C. asteris* exhibited a growth both on packings and in the liquid phase. For these cultures, the dry matters of each phase were weighted separately.

2.6. Support recycling procedure

Different washing procedures of the stainless steel packing material were carried out. Inoculation was performed with the above mentioned spore inoculum. Cultures were maintained for 49 days. The treatment A consisted to rinse the support with cold distilled water, while in treatment B the support was immersed for 24 h in NaOH solution (20%) before being rinsed with distilled water until all NaOH was removed (estimated by the neutral pH of the last washing water). Treatment C is the same as treatment B with an additional step of immersion of the support in distilled water at 100 °C for 1 h.

2.7. Cultures in bioreactors

C. asteris growth and astin C production as a function of culture medium pH value was investigated in 400 mL bioreactor (DASGIP, Jülich, Germany) containing 220 mL of MEAc medium and fitted with a stainless steel structured packing. Cultures were performed for 36 days with or without pH regulation. In case of its regulation the set point was fixed at 4.8. Bioreactors were not mechanically stirred, only aerated (at 58.8 and 117.6 mL/min). The development of *C. asteris* was assessed through visual observations and the final astin C concentration was quantified after 36 days.

2.8. Astin C quantification

Astin C measurements were initially performed through HPLC-ESI-MS analysis. Concerning the media optimization experiments, supernatant obtained from 11 day-old cultures were treated with ethyl acetate (1:1 volume ratio), evaporated, and resolved in 100 μ L methanol (50%). Analyses were performed on an HP1090/XCT6330 system (Agilent Technologies, Waldbron, Germany). The astin C amount, deduced from the integral of the 7.7 min-peak in the extracted ion

chromatogram at m/z 570.2 in positive mode, was expressed in percentage compared to samples obtained in control conditions with unaltered medium composition. An alternative and faster astin C quantification method based on HPLC-UV has been further developed. This method was subsequently employed to quantify astin C, in each sample analyzed, on the basis of a calibration curve obtained with of astin C standard of known concentrations. Samples were first centrifuged (30 800 x g, 5 min) and 500 μ L of supernatant were mixed with 500 μ L methanol. The solutions were then agitated for 30 min and filtered on cellulose acetate (0.22 μ m) before analysis. The quantification was done while injecting 100 μ L of solution in an Agilent 1100 SERIES HPLC (Agilent Technologies) coupled to a UV-VIS detector (detection at 215 nm). The column was an Agilent Zorbax C18 Eclipse Plus (3x150 mm 5 μ M). The solvents were (A) MilliQ water with 0.1% trifluoroacetic acid and (B) acetonitrile (HPLC grade) with 0.1% trifluoroacetic acid (solvent flow rate: 1 mL/min). The solvent gradient was set as following: start – 75% A / 25% B; from 0 to 6 min – from 75% A / 25% B to 70% A / 30% B; from 6 to 8 min – from 70% A / 30% B to 0% A / 100% B; from 8 to 13 min – 0% A / 100% B; from 13 to 15 min – from 0% A / 100% B to 75% A / 25% B; from 15 to 20 min – 75% A / 25% B. The elution time of astin C standard under the mentioned above conditions was at 6.30 min.

2.9. Scanning electron microscopy (SEM)

Samples from two-week old planktonic culture (PL) and 10 x 10 mm square cuts from ICCs on steel grids were first fixed by immersion in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer pH 7.4 for 24 h, and then rinsed three times in 0.2 M buffer solution. Pellets from planktonic cultures and biofilms on steel grid were dehydrated through an ethanol series up to 100% followed by critical point drying using CO₂. The dried samples were put on glass slides covered with double-side carbon tape. In order to look at both sides of the support grid, the square samples were mounted in standing position. Detached parts of biofilms were put aside either upside up or upside down to see their outer surface as well as their contact with the support. The grid surfaces from where biofilms were detached were also examined. All samples were 20 nm Pt-coated in a Balzers SCD-030 sputtering unit (Balzers Union Ltd., Balzers, Liechtenstein) before being observed in an ESEM FEI QUANTA 600 under 15-20 kV accelerating voltage, high vacuum conditions and using an Everhart-Thornley -secondary electron detector.

3. Results

3.1. Optimization of the culture medium

A culture medium derived from MEA (deemed to be appropriate for fungi growth [19,20]), with 10 g/L of malt extract, was used to investigate the effect of different carbon sources (glucose, mannose, and sucrose) at two concentrations on *C. asteris* growth and astin C production. When mannose was used in the culture medium, both biomass and astin C production increased compared to media that contained glucose or sucrose as sugar source (**Figure 1A**).

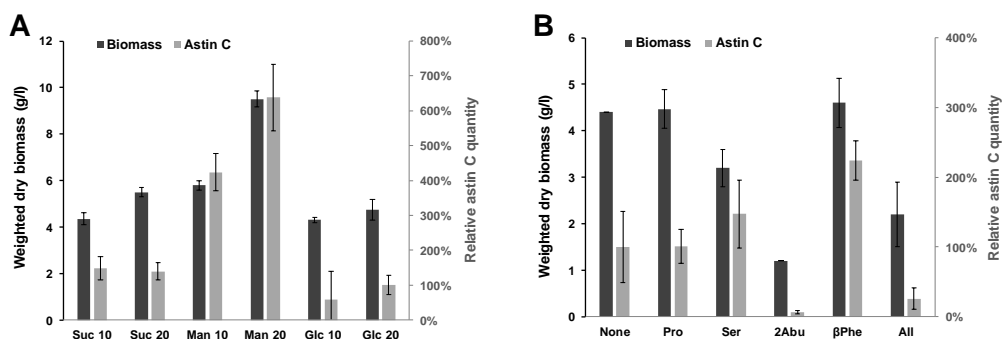


Figure 1: Effect of medium modifications on *C. asteris* biomass and astin C production. Cultivations were performed at 22 °C in planktonic flasks under shaking at 150 rpm. Mean values and standard deviations were calculated from the data obtained from triplicate flasks filled with 50 mL of one of the following medium: **A.** Culture medium (malt extract 10 g/L, meat peptone 1 g/L, trace elements) containing either 10 or 20 g/L of one of the sugars: sucrose (Suc), mannose (Man), or glucose (Glc). Relative astin C quantity was expressed in percentage of the value obtained with 20 g/L of glucose (control condition). **B.** Culture medium (malt extract 10 g/L, glucose 10 g/L, meat peptone 1 g/L, trace elements) supplemented with the astin C building blocks: L-proline (Pro) 0.5 g/L, L-serine (Ser) 0.5 g/L, L-2-aminobutyric acid (2Abu) 1 g/L, L- β -phenylalanine (β -Phe) 0.5 g/L, or a combination of all four amino acids. Relative astin C quantity was expressed in percentage of the value obtained in the unsupplemented medium (control condition).

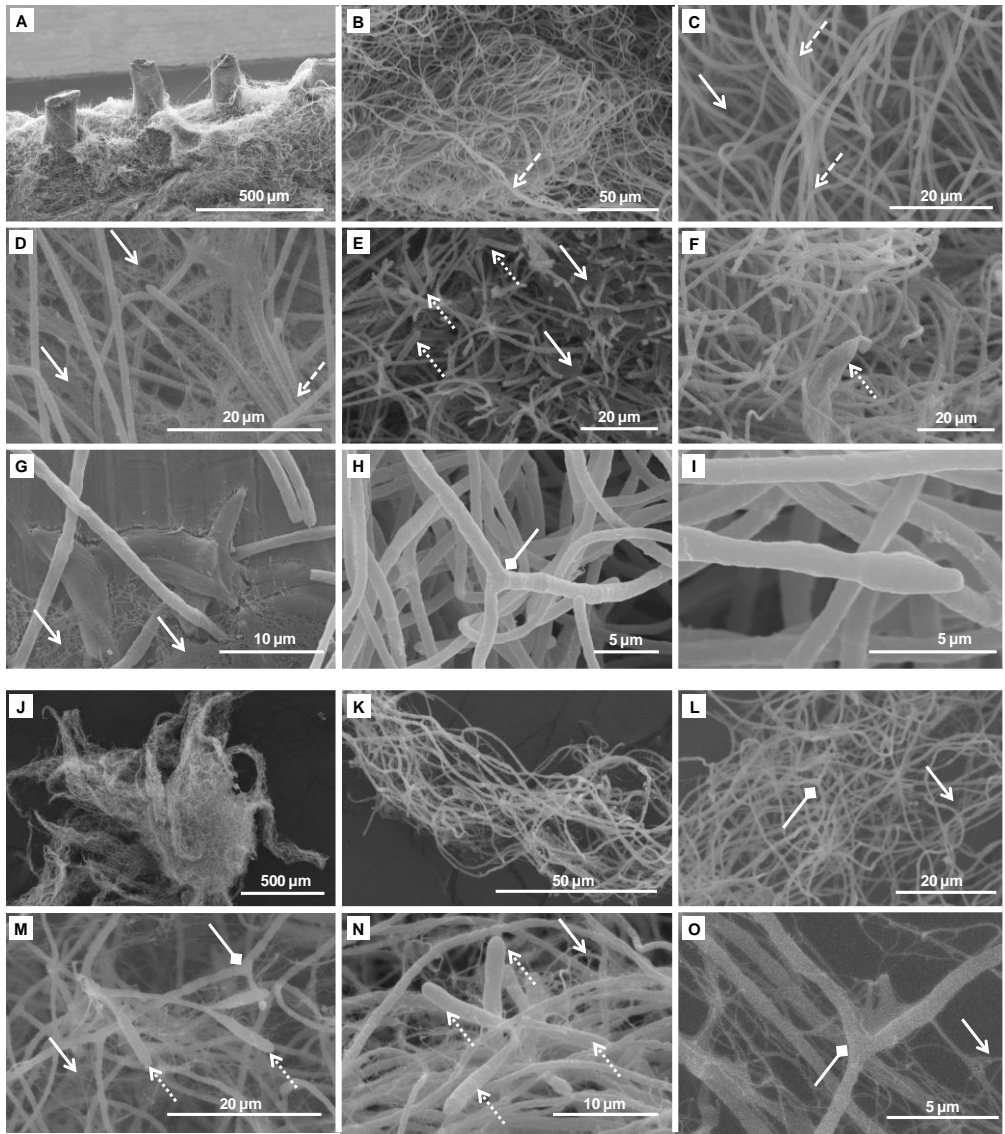
In order to further increase production of both fungal biomass and astin C, we investigated the effect of the addition of selected amino acids that represent building blocks of astin C into the culture medium. The addition of only β -Phe resulted in an increased formation of astin C (**Figure 1B**). Interestingly, addition of L-2-aminobutyric acid had a negative effect on both biomass and astin C. Due to the positive effects of mannose and β -Phe on astin C production, an adapted medium containing these ingredients, named MEAb, was formulated (see section 2 for details). This culture medium was used in subsequent experiments.

3.2. Comparison between two cultivation methods of *C. asteris*: planktonic and immobilized-cell conditions

3.2.1. Macroscopic observations and characterization of the fungus through electron microscopy

C. asteris was grown in PL and ICCs during two weeks in the MEAb medium. Macroscopic observations of the fungus development showed a high variability of the morphology depending on the culture conditions. *C. asteris* has grown as a suspension of small pellets (diameter of 2 to 5 mm) in planktonic conditions. On the contrary, the fungus has developed as a dense biofilm on the stainless steel support. In the formed biofilm, two distinct layers could be distinguished: a hydrophilic lower layer in contact with the aqueous medium with a dark pink color and an upper layer in contact mainly with air, displaying a light pink color with high contact angle droplets suggesting hydrophobic properties.

Cell morphology in biofilm (**Figure 2A-I**) and pellet (**Figure 2J-O**) was further characterized by SEM.



◁ **Figure 2:** Scanning electron microscopy (SEM)-views of *C. asteris* grown during 14 days in the MEAb medium (25 °C-160 rpm), as biofilm on the steel support (**A-I**) and as pellets in planktonic cultures (**J-O**). The following elements are indicated by a dedicated arrow: local bundle arrangements in microfibrils (dashed arrow), filamentous extracellular polymeric substances (solid arrow), spore-forming structures (dotted arrow), and hyphae showing branching (lozenge shaped arrow head). **A.** General views showing grid bars that have been sectioned and are covered by massive strands of filaments. **B, C.** Views of the external part of the biofilm. **D.** View of the middle part of the biofilm. **E, F, G.** Views of the deepest part of the biofilm attached to the steel grid support. **H, I.** Details of the mycelium filaments (hyphae) showing branching and terminal ends. **J.** General view of a star-shaped pellet (a central core and several arms) formed in PLs. **K.** View of the middle part of an arm formed by a bundle of roughly parallel filaments. **L.** View of the middle part of arms. **M.** Details of the arm filaments with branching hyphae, local irregular cellular patterns and frequent characteristic fructifications in the middle part of arms. Spores exhibit a characteristic round-shaped terminal end and a constricted junction with the mother-hyphae. **N.** View of the central core pellet with a regular terminal end, terminal spores and abundant EPS. **O.** Detailed view of the hyphae showing branching and EPS interlacing the hyphae.

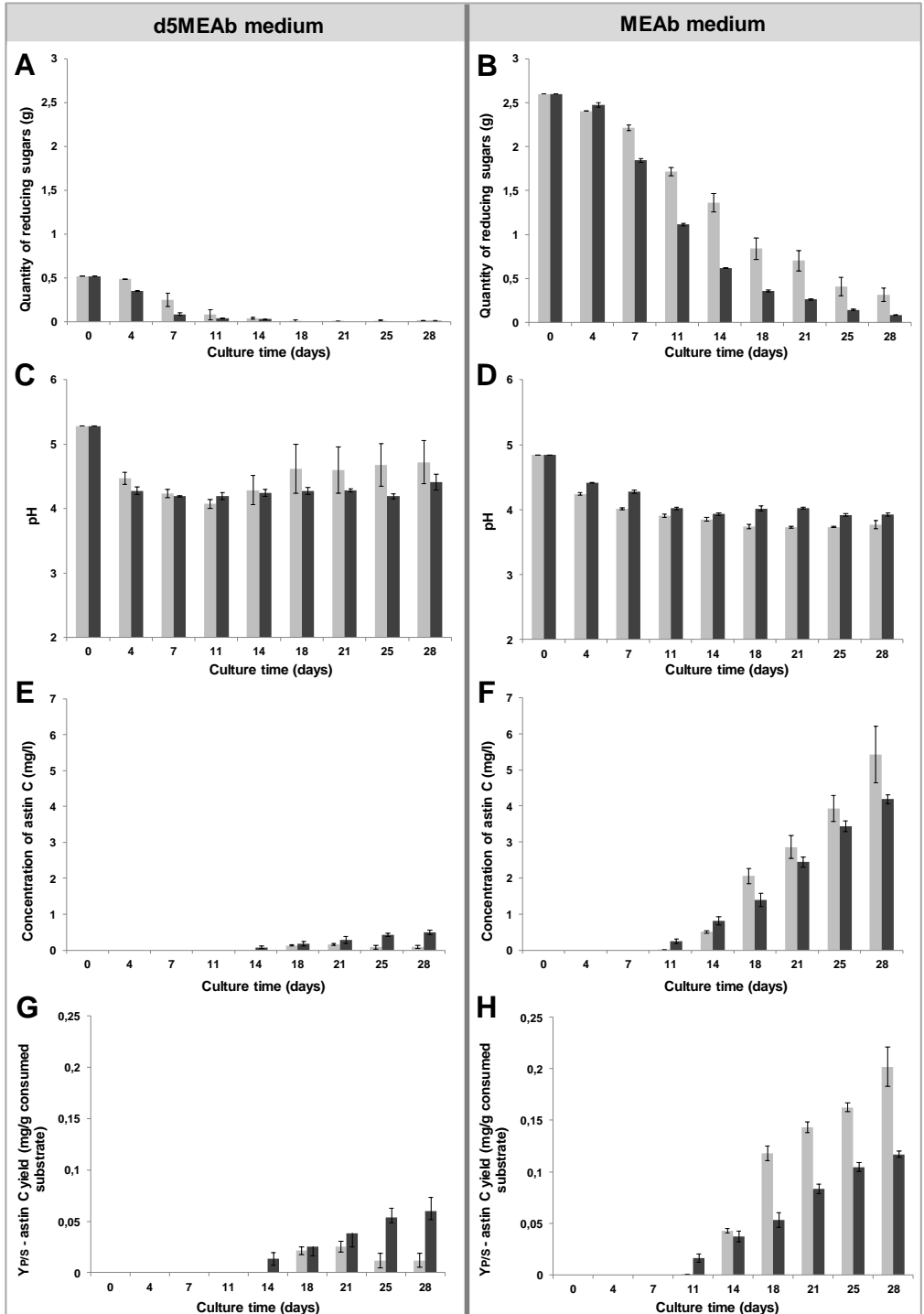
The overview of the biofilm form on the solid support is depicted in **Figure 2A**. The biofilm appeared externally as a very dense framework of agglomerated hyphae covering the steel stems of the packing. The fungus also grew through the holes of the packing. The biofilm aspect at the contact surface between the fungus and the steel support can be observed in **Figure 2B-I**. Two layers can be distinguished on the biofilm pictures with a superior layer composed of fungal hyphae, not in direct contact with the steel support, and an inferior layer containing spores. Spores appeared dark grey because they are fixed to the metallic support (**Figure 2 E, G**). On the contrary, hyphae appeared light grey because they were not in direct contact with the metallic support (**Figure 2 B-D, H**). On the surface of the biofilm, hyphae tended to form microfibrils (**Figure 2 B, C**, dashed arrows). Extracellular Polymeric Substances (EPS) could be seen in the structure, forming a dense framework between fungal cells. The density of EPS framework tended to increase with the distance from the biofilm surface (**Figure 2 C-E, G**, white arrows). The deepest part of the biofilm was rich in germinated spores (dotted arrows), and in EPS organized in complex 3-D structures (**Figure 2 E-G**). Both in biofilm and pellet forms, hyphae had a thin diameter of about 1 μm (**Figure 2I**) and displayed some rare ramifications (**Figure 2 H, L, O**, lozenge shaped arrow head). The structure of fungal pellets was much less dense. Fungal pellets consisted of a core zone connected to multiple branches (**Figure 2J**). The density of EPS tended to decrease from the core to the external parts. However, the EPS framework was much less thick in pellets than in biofilms (**Figure 2 L-O**, white arrows). The EPS in the pellets were organized in the form of diffuse links between hyphae. In biofilms, they were rather clustered in dense sheets linking fungal cells together. Extremities of the pellet arms were composed of one or two hyphae with a surface covered by EPS chains, but too short to allow connections. A higher hyphae density seems to induce the clustering of the fungus by the EPS as in the core zone (**Figure 2 M, N**, white arrows). An interesting observation is that fructifying parts could be pointed out both in central and distal parts of the pellets (**Figure 2 M, N**,

dotted arrows). The density of spore-forming structures was quite variable based on the individual pellet which was considered, probably depending on the growth state. These spores showed a conical shape and were formed at hyphal extremities. A common sequence of events occurred regularly at hyphal extremities: one spore was formed by the end of the hyphal and was followed with two other symmetric spores at the end of the structure.

3.2.2. Biomass and astin C production by *C. asteris* in planktonic and ICC conditions

C. asteris growth and astin C production was monitored during 4 weeks, in PL and ICCs, through the determination of the final dry biomass, the quantity of reducing sugars, the pH value, and the astin C quantification. Experiments were performed in the MEAb medium and in the d5MEAb with the purpose of setting up a production process based on a cheaper culture. As shown in **Figure 3A**, all reducing sugars from the d5MEAb medium were consumed within 14 days in both culture conditions. In the MEAb medium (**Figure 3B**), reducing sugar concentrations in PL samples were always lower than in ICC samples after 7 days of cultivation (significant difference with a p -value < 0.05). In both conditions, a remainder of reducing sugars persisted at the end of the 4 culture-weeks, but the consumption was slowed down in ICCs.

The pH evolution of the medium in PL and ICCs was also investigated in the d5MEAb (**Figure 3C**) and the MEAb medium (**Figure 3D**). The pH evolution was different according to both medium concentration and culture mode. The d5MEAb medium showed the broadest pH variation whatever culture type considered. Indeed, the pH value first decreased during the first two weeks and then increased until the end of cultivation, suggesting a switch in the fungal metabolism. This effect is even more marked in ICCs. At the end of cultivation, the highest pH value (4.7) was measured with ICCs in the d5MEAb. The medium with the lowest concentration of reducing sugars yielded to the highest pH increase. In comparison, the pH values recorded in the MEAb medium did not display such a pH increase.



◀ **Figure 3:** Comparison between two cultivation methods of *C. asteris*: ICCs (light grey bars) and PL (dark grey bars) in the MEAb or the d5MEAb. Cultivations were performed at 25 °C in flasks under shaking at 160 rpm. Mean values and standard deviations were calculated from the data obtained from triplicate flasks filled with 100 mL of culture medium. **A, B.** Follow-up of the reducing sugars consumption, quantified via the DNS method, in ICCs and PLs in **(A)** d5MEAb medium or **(B)** MEAb medium. **C, D.** Follow-up of the pH evolution of **(C)** d5MEAb medium or **(D)** MEAb medium while growing *C. asteris* in ICCs and PLs. **E, F.** Follow-up of the astin C concentration (mg/mL) in the ICCs and PLs of *C. asteris* in **(E)** d5MEAb medium or **(F)** MEAb medium. **G, H.** Mean value of astin C yield ($Y_{P/S}$) in mg per gram of consumed reducing sugar in time, with ICC and PL conditions, in **(G)** d5MEAb medium or **(H)** MEAb medium.

After 28 days, cell growth was estimated by dry biomass measurement, enabling thus the determination of corresponding biomass yields, taking the residual substrate into account. As shown in **Table 1**, no significant difference was found between biomasses generated in the d5MEAb culture medium in PLs and ICCs. However, there was a significant difference of dry biomass in the MEAb medium (p -value < 0.05), where the biomass developed in ICCs was 39% higher than in PL condition at the end of the cultivation period. The biomass yields obtained in the MEAb medium were also almost twice higher in ICCs compared to PLs cultures.

Table 1: Mean values of the final weighted dry biomass, the biomass yield $Y_{X/S}$, and the specific astin C production $Y_{P/X}$ in planktonic and immobilized-cell conditions.

		Planktonic conditions		Immobilized-cell conditions	
Weighted dry biomass^{a)}					
(in g/L)					
<i>Medium</i>	d5MEAb	1.9	(± 0.4)	1.9	(± 0.1)
	MEAb	5.6	(± 0.3)	9.2	(± 0.3)
Biomass yield - $Y_{X/S}$^{b)}					
(in g/g consumed sugars)					
<i>Medium</i>	d5MEAb	0.36	(± 0.07)	0.37	(± 0.02)
	MEAb	0.22	(± 0.01)	0.40	(± 0.003)
Specific astin C production					
$Y_{P/X}$^{c)} (in mg/g of dry biomass)					
<i>Medium</i>	d5MEAb	0.27	(± 0.06)	0.05	(± 0.03)
	MEAb	0.76	(± 0.07)	0.59	(± 0.07)

^{a)}Mean values and standard deviations were calculated from the data obtained at the end of the 28 culture-days in flasks filled with 100 mL medium, done in triplicate ;

^{b)} $Y_{X/S}$ = Final dry biomass / (initial reducing sugars – final reducing sugars), calculated for each single flask, mean values and standard deviations were calculated from the $Y_{X/S}$ of the triplicate flasks ;

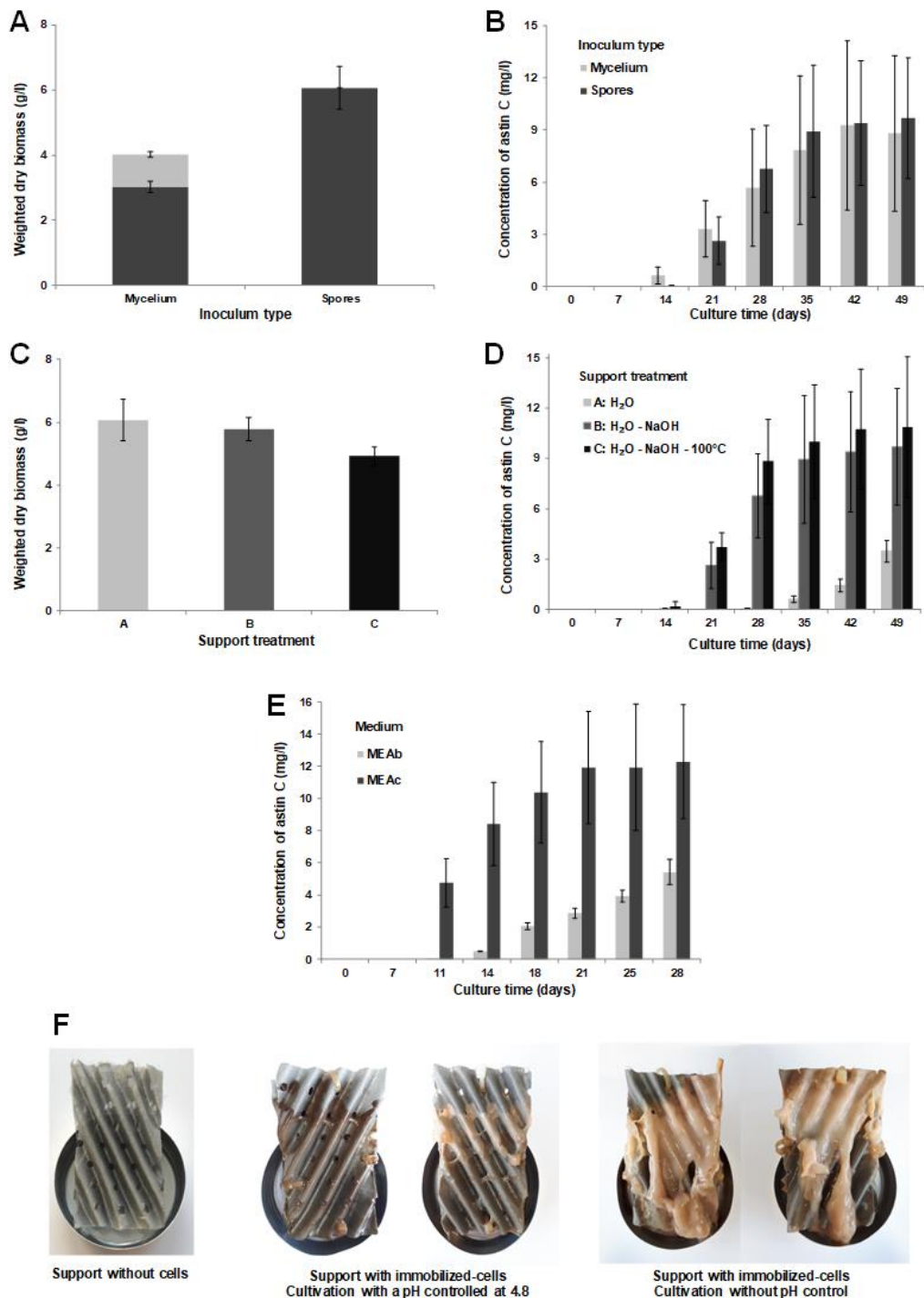
^{c)} $Y_{P/X}$ = Final quantity of astin C / Final dry biomass; calculated for each single flask, mean values and standard deviations were calculated from the $Y_{P/X}$ of the triplicate flasks.

In those experimental conditions, astin C production was monitored over time. As shown in **Figure 3E**, and **Figure 3F**, astin C detection in the culture medium started between 11 and 14 days whatever the culture medium and cultivation conditions. Moreover, astin C production was higher in the concentrated culture medium (i.e. MEAb). For this experimental condition, the final astin C production was increased by 30% in ICCs compared to planktonic condition. In order to highlight any correlation between astin C production and cell concentration, specific production (i.e. $Y_{P/X}$, amount of astin C produced per g of biomass) was calculated (**Table 1**). Higher astin production per fungal dry mass unit was obtained in the MEAb in PL as well as ICC conditions. In addition, the fungus grown in planktonic cultures displayed a higher specific astin C production per mass unit whatever the medium considered. Nonetheless, this variation between PL and ICCs was more prominent in the diluted medium. As the higher specific production per mass unit was in PLs, the 30% higher astin C production observed in ICCs (**Figure 3 E, F**) is more likely related to the enhanced biomass development observed on stainless steel packings. Yield of astin C (i.e. $Y_{P/S}$, amount of astin C produced per g of reducing sugar consumed) was evaluated considering cost-efficiency (**Figure 3 G, H**). $Y_{P/S}$ was higher in the MEAb medium for both ICCs and PLs. Therefore, the use of MEAb was, in addition to be favorable to astin C production, also more cost-effective than the use of d5MEAb. Secondly, while focusing on the MEAb medium, the $Y_{P/S}$ of astin C was higher in ICCs than in PLs after 14 days of cultivation. Consequently the combination of ICCs with the MEAb medium appeared to be the most appropriate for the production of astin C by *C. asteris*. Therefore, we intensified our efforts to develop a suitable astin C production system based on these culture conditions.

3.3. Optimization of astin C production by C. asteris in immobilized-cell cultures

3.3.1. Effect of the inoculum type on immobilized cell cultures of *C. asteris*

The effect of the inoculum type, either with 1 mL of 15 or 55 day-old pre-cultures, on cell growth and astin C production was investigated. The inoculum from the 15 day-old culture containing mainly mycelium structures was designated as mycelium inoculum, while the one from the 55 day-old culture, enriched with spores, was nominated as spore inoculum. Cultures were inoculated with the two types of inoculum and grown in the MEAb medium under immobilized-cell condition during 49 days. Biomass was determined at the end of the culture (i.e. after 49 days, **Figure 4A**) while astin C production was monitored over time (**Figure 4B**).



◀ **Figure 4:** Evaluation, in ICCs, of the effect of some cultivation parameters on the production of astin C and the development of *C. asteris*. **A. B.** Influence of the inoculum type on astin C production and *C. asteris* growth. Cultivations were performed under shaking at 160 rpm, at 25 °C, in flasks filled with 100 mL of the MEAb medium. Mean values and standard deviations were calculated from the data obtained from triplicate flasks inoculated either with 1 mL of hyphae from a 15 day-old culture (mycelium) or 1 mL of a mixture of hyphae and spores obtained from a 55 day-old culture (spores). **A.** The fungal final dry biomass (g/L) attached to the support (dark grey) or released into the medium (light grey). **B.** The follow-up of the astin C concentration (mg/mL). **C. D.** Evaluation of the impact of the washing treatment applied to the support on *C. asteris* attachment and development (**C**) and on astin C production (**D**). The supports used to immobilize the cells were previously washed either with a treatment A (H₂O only), a treatment B (H₂O followed by 24 h immersion in NaOH 20%), or a treatment C (identical to treatment B with an additional step of 1 h immersion in 100 °C H₂O). Cultivations were performed under shaking at 160 rpm, at 25 °C, in flasks filled with 100 mL of the MEAb medium inoculated with spores. Mean values and standard deviations were calculated from the data obtained from triplicate flasks. **E.** Influence of the medium composition on the astin C production in ICCs. Cultivations were performed under shaking at 160 rpm, at 25 °C, in flasks filled with 100 mL of culture medium inoculated with spores. Mean values and standard deviations were calculated from the data obtained from triplicate flasks filled either with the MEAb or MEAc culture medium. **F.** Evaluation of the development of *C. asteris* on the stainless steel support in two different pH cultivation conditions. *C. asteris* was grown as a biofilm during 36 days at 25 °C in immobilized-cell condition, in DASGIP bioreactors with a pH controlled at 4.8 or without any pH control. The visual aspects of the stainless steel supports obtained after cultivation of *C. asteris* with the pH controlled at 4.8, and without the pH control, are compared with the visual aspect of the support without cells.

After 49 culture-days, *C. asteris* from the spore inoculum grew exclusively as a biofilm on the solid support. However, in *C. asteris* cultures inoculated from mycelium inoculum, a development of fungi hyphae, in pellets form, in the liquid medium appeared after 7 culture-days. Nonetheless, at the end of the cultivation period, 75% of the biomass was attached to the support in these cultures. As shown on **Figure 4A**, spore inoculum led to higher biomass in comparison with the mycelium inoculum. Astin C appeared in the culture medium from both inocula from day 14. Taken together, these data indicate that the inoculation with spores, even if it does not provide a significant improvement in astin C production, is favorable to the development of the fungal biomass and its attachment to the support. Cells growing exclusively as a biofilm on the support are of great advantage while considering further purification and recovery of the astin C secreted into the medium.

3.3.2. Implementation of a production methodology: support recycling issues

The question on the re-use of the steel packing is of prime importance with a view on the implementation of a routine production methodology. Based on our predefined production system (ICCs in the MEAb with spore inoculum), we investigated the effect of three different washing procedures of the packing on the fungal biomass development (**Figure 4C**) and on the astin C production (**Figure 4D**). Three treatments (A, B or C) of varying intensity, as described in Section 2, were applied to

prepare supports. Treatment A encompassed only a step of distilled water rinse, while a preliminary step of immersion in NaOH solution (20% - 24 h) was applied in treatment B, as well as for treatment C which further comprised a final 1h immersion in 100 °C distilled water. Two groups with significantly different astin C production were found for all quantifications (except for culture-day 14). The first group involved support treatment A (with the lowest astin C concentrations), while the second group constituted support treatments B and C (highest astin C production). The pre-washing step with NaOH (20%, 24 h) was mandatory to reach higher amounts of astin C, whereas the boiled water treatment did not significantly improve the production.

3.3.3. Optimization of immobilized-cell culture conditions to reach higher astin C yields

In a last set of experiments, *C. asteris* was grown in immobilized-cell condition in another culture medium (MEAc) with different additional carbon sources such as glucose and a greater global abundance of reducing sugars than MEAb, but without β -phenylalanine. The purpose of this experiment was to evaluate the growth of *C. asteris* and the astin C production within a culture medium devoid of its very economically costly astin C building block. *C. asteris* was inoculated using spores in MEAb and MEAc, in the ICC system, and the astin C production was monitored up to 28 days (**Figure 4E**). A large increase in the production of astin C was detected in the MEAc in comparison with the MEAb medium. After four weeks, the final astin C amount was 2.3 times higher in the MEAc medium. The final dry biomass obtained was also slightly enhanced compared to the MEAb medium (13.5 and 9.2 g/L respectively). The corresponding specific astin C production Y_{PX} was 0.89 mg/g of fungal mass unit. With our defined conditions, the MEAc was a more appropriate culture medium than the MEAb to reach higher production yield of astin C.

In order to further optimize the culture conditions, we also investigated the impact of the pH control on *C. asteris* development and astin C production. We set up a cultivation experiment in multiple DASGIP system bioreactors. *C. asteris* was grown on support, in the MEAc culture medium, at two different pH conditions for 36 days. Two bioreactors were performed with a pH controlled at 4.8, and two others were left without any pH control. At the end of the cultivation the pH of the two latter cultures decreased to 3.5. The development of the fungus was assessed with a visual observation at the end of cultivation. Bioreactors run at pH 4.8 displayed a very weak development of *C. asteris* (**Figure 4F**). In contrast, the fungus was able to develop well on the entire support in bioreactors without any pH control. Astin C was quantified at the end of cultivation for indicative purpose. Production of astin C was only detected in the bioreactors without pH control (final concentration of 0.50 ± 0.03 mg/mL). No production of astin C was detected in the pH controlled bioreactors. Therefore, the development of the fungus was associated with medium acidification, which is likely also favorable in the matter of astin C production and/or stability.

4. Discussion

The aim of this study was to develop and to optimize a cultivation system for the recently isolated *C. asteris* fungus in order to enhance the production rate of astin C. The first step consisted in optimizing the composition of the culture medium. Mannose, one of the sugars considered, appears frequently in form of its polymer mannan in plants [22]. The beneficial effect of mannose observed in *C. asteris* could be correlated to its endophytic lifestyle, which probably has resulted in an adapted utilization of plant substances. During astin C biosynthesis, the formation of β -phenylalanine seems to constitute a bottleneck reaction since feeding of that building block had a positive effect on astin C production. Nonetheless, the lack of β -phenylalanine in the medium was also offset while providing additional carbon sources in larger amount. The feeding with 2-aminobutyric acid, another astin C building block, inhibited the astin C production and the *C. asteris* growth. The inhibitory effect of aminobutyric acid-type compounds on the development of fungal species has already been reported [23].

As advantages of ICCs were previously demonstrated concerning enzymes or secondary metabolites production by filamentous fungi [14,24], this approach was investigated to grow *C. asteris*. The thin width of its fungal hyphae ($\sim 1 \mu\text{m}$) compared to *Aspergillus* spp for example ($\sim 3\text{-}6 \mu\text{m}$) [25], was likely to contribute positively to its development through the holes of the steel support in ICCs. In the MEAb medium, the variations of *C. asteris* metabolism (in terms of substrate consumption, cells development and astin C production) could be correlated to the different mycelium morphologies observed either in immobilized-cell or planktonic conditions. The structure of the biofilm grown on support was much more developed than in pellets formed in planktonic conditions. In biofilms, the access to substrates is made difficult by the thickness of the layers which can explain the slower sugar consumption in ICCs compared to PLs [26]. However, substrate consumption was twice more efficient in term of biomass development in ICCs than in PLs ($Y_{X/S}$ of 0.40 and 0.22 g/g respectively). An enhanced ribosome production and protein translation efficiency in cell immobilized as biofilm have already been reported in previous studies [26,27]. Taken together with the higher complexity of the extracellular matrix in ICCs than in PL, these results suggest a higher level of biosynthesis of immobilized-cells. However, it must be kept in mind that the dry mass measured in biofilms also includes the matrix, which is much more complex than in PLs, as depicted on the SEM views.

In terms of astin C production, despite a higher specific astin C production per fungal mass unit ($Y_{P/X}$) in planktonic conditions, the final product titers were 30% higher in immobilized-cell conditions. As only little information is available about *C. asteris*, it is difficult to explain the higher production rates obtained in ICCs [13,16]. However, it is likely that the enhanced astin C production in ICCs was mainly due to the 39% higher biomass development in biofilm form. This observation also means that astin C does not inhibit the growth of the fungus in this concentration range.

In two different set of experiments, pH was highlighted as a key parameter impacting *C. asteris* growth and astin C production. On one hand, in the d5MEAb

medium a negative impact of the pH alkalisation on biomass development and astin C production was observed. On the other hand, similar results were obtained when, in bioreactor assays, the pH was controlled at 4.8.

With the purpose of optimizing the cultivation of *C. asteris* as immobilized-cells, some parameters likely to affect the fungus-support interactions were investigated. Firstly, the inoculation type with spores rather than with mycelium had shown a positive impact in terms of biomass attachment and development onto the support. The promoted attachment of spores was further confirmed on SEM views, highlighting the presence onto the steel support of a layer composed of spores supporting a thick layer of hyphae. Although the case of *C. asteris* is not documented, structural differences in spores and hyphae have been investigated previously in other filamentous fungi such as *Aspergillus* spp. or *Neurospora crassa* [28,29]. The cell wall of spores is surrounded by a melanin coat, itself coated with a rodlet layer composed by a monolayer of hydrophobins forming fibrils. Hydrophobin monolayers are really robust and can reverse the wettability of the surface on which they are formed, allowing spores to adapt to any type of surface [30]. This type of structure could explain this ability of attachment observed with the spores of *C. asteris*. Secondly, the impact of the washing treatment applied to the steel support on the astin C production and the fungus development was investigated. Although the efficiency of the cleaning procedure to remove fungal biofilm is not well documented, some bacterial biofilm cleaning procedures have been reported. The use of NaOH as a caustic cleaner is the most common (and aggressive) option to remove bacterial biofilm [31–33], as well as EPS [34], from stainless steel support. The ability of *C. asteris* spores to attach to the treated support was not affected, unlike the production of astin C which was very poor without NaOH treatment. Therefore, the washing treatment applied leads to a modification of the physicochemical properties of the support causing a still inexplicable variation of the fungal productivity.

The specific astin C production obtained at best in this study was 0.89 mg/g of the fungal mass unit. For comparison purpose, the specific astin C titer extracted from *Aster tataricus* roots was estimated at 0.16 mg/g of plant materials (8 g of astin C from 50 kg of roots) [11]. This study represents a step forward in the astin C production process. Nonetheless, taking into account that fermentation of 28 days, in 100 mL of rich media, is required to produce 1.23 mg of astin C with *C. asteris* (i.e. an equivalent astin C amount obtained from 7.7 g of *A. tataricus* roots), the titers obtained are far to meet industrial yield requirements.

Whether this contribution is not tremendous regarding astin C production, it enables to considerably facilitate the recovery of the product. Multiple laborious steps are required to extract astin C from *A. tataricus*, unlike with ICCs of *C. asteris* growing exclusively on the support and not in the culture medium, enabling a facilitate recovery and purification of the secreted astin C. On the whole, the development of ICCs of *C. asteris* has already shown promising improvement in the astin C production associate with a decrease in the downstream processing cost. Further optimizations are required to enable astin C to take the path followed decades ago by the penicillin initially produced also in poor titers by *Penicillium chrysogenum* [35].

- **Acknowledgement**

This work was funded by the ERA-IB 2 program (ASTINPROD project, ERA-IB-15-039) and the international laboratory Microbial Secondary Metabolites MOM (University of Lille/Gembloux Agro-Bio Tech University of Liege).

- **Conflict of interest**

The authors declare no financial or commercial conflict of interest.

5. References

1. Kusari, S., et al. (2012) Chemical ecology of endophytic fungi: origins of secondary metabolites. *Chem. Biol.*, **19** (7), 792–798.
2. Masi, M., et al. (2018) Bioactive Metabolites from Pathogenic and Endophytic Fungi of Forest Trees. *Curr. Med. Chem.*, **25** (2), 208–252.
3. Jahn, L., et al. (2017) *Cyanodermella asteris* sp. nov. (Ostropales) from the inflorescence axis of *Aster tataricus*. *Mycotaxon Issn*, **132** (March), 107–123.
4. Zhang, Y., et al. (2012) Inhibition of human gastric carcinoma cell growth in vitro by a polysaccharide from *Aster tataricus*. *Int. J. Biol. Macromol.*, **51** (4), 509–513.
5. Yu, P., et al. (2015) Expectorant, antitussive, anti-inflammatory activities and compositional analysis of *Aster tataricus*. *J. Ethnopharmacol.*, **164**, 328–333.
6. Jahn, L., et al. (2017) Linking secondary metabolites to biosynthesis genes in the fungal endophyte *Cyanodermella asteris*: The anti-cancer bisanthraquinone skyrin. *J. Biotechnol.*, **257** (June), 233–239.
7. Thomas, S., et al. (2019) Production of antitumor astins by the endophyte *Cyanodermella asteris* isolated from the medicinal plant *Aster tataricus*. *Submitted*.
8. Kosemura, S., et al. (1993) Isolation and Structure of Asterin, a new halogenated cyclic pentapeptide from *Aster tataricus*. *Tetrahedron Lett.*, **34** (8), 1291–1294.
9. Saviano, G., et al. (2004) Influence of conformational flexibility on biological activity in cyclic astin analogues. *Biopolym. - Pept. Sci. Sect.*, **76** (6), 477–484.
10. Rossi, F., et al. (2004) New antitumour cyclic astin analogues: Synthesis, conformation and bioactivity. *J. Pept. Sci.*, **10** (2), 92–102.
11. Shen, Y., 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** (3), 260–268.
12. Li, F., et al. (2018) Design and synthesis of plant cyclopeptide Astin C analogues and investigation of their immunosuppressive activity. *Bioorganic Med. Chem. Lett.*, **28** (14), 2523–2527.
13. Vandermies, M., et al. (2018) *Yarrowia lipolytica* morphological mutant enables lasting in situ immobilization in bioreactor. 5473–5482.
14. Zune, Q., et al. (2015) A fungal biofilm reactor based on metal structured packing improves the quality of a *Gla*::GFP fusion protein produced by *Aspergillus oryzae*. 6241–6254.
15. Papagianni, M., and Matthey, M. (2004) Physiological aspects of free and immobilized *Aspergillus niger* cultures producing citric acid under various glucose concentrations. *Process Biochem.*, **39** (12), 1963–1970.
16. Mizunuma, T., et al. (2007) A mycelium with polyelectrolyte complex-bunched hyphae: preparation and fermentation performance. *Colloids Surf. B. Biointerfaces*, **56** (1–2), 155–160.
17. Villena, G.K., and Gutiérrez-Correa, M. (2007) Morphological patterns of *Aspergillus niger* biofilms and pellets related to lignocellulolytic enzyme productivities. *Lett. Appl. Microbiol.*, **45** (3), 231–237.
18. Weenink, X.O., et al. (2006) A new method for screening and isolation of hypersecretion mutants in *Aspergillus niger*. *Appl. Microbiol. Biotechnol.*, **69** (6), 711–717.
19. Yazid, S.N.E., et al. (2018) Formulation of maize- and peanut-based semi-synthetic growth media for the ecophysiological studies of aflatoxigenic *Aspergillus flavus* in maize and peanut agro-ecosystems. *Int. J. Food Microbiol.*, **282**, 57–65.
20. Weissfeld, A.S., et al. (2013) Optimal Media for Use in Air Sampling To Detect Cultivable Bacteria and Fungi in the Pharmacy. *J. Clin. Microbiol.*, **51** (10), 3172–3175.

21. Miller, G. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.*, **31** (3), 426–428.
22. Melton, L.D., et al. (2009) Mannans in primary and secondary plant cell walls. *New Zeal. J. For. Sci.*, **39**.
23. Fu, D., et al. (2017) Comparison of the effects of three types of aminobutyric acids on the control of *Penicillium expansum* infection in pear fruit. *J. Sci. Food Agric.*, **97** (5), 1497–1501.
24. Baños, J.G., et al. (2009) High lovastatin production by *Aspergillus terreus* in solid-state fermentation on polyurethane foam: An artificial inert support. *J. Biosci. Bioeng.*, **108** (2), 105–110.
25. Shah, A.A., et al. (2013) Diagnostic Accuracy of Histopathologic and Cytopathologic Examination of *Aspergillus* Species. 55–61.
26. Fanning, S., and Mitchell, A.P. (2012) Fungal biofilms. *PLoS Pathog.*, **8** (4), 1–4.
27. Gutiérrez-Correa, M., et al. (2012) Recent advances on filamentous fungal biofilms for industrial uses. *Appl. Biochem. Biotechnol.*, **167** (5), 1235–1253.
28. Beauvais, A., et al. (2014) *Aspergillus* cell wall and biofilm. *Mycopathologia*, **178** (5–6), 371–377.
29. Macindoe, I., et al. (2011) Monolayers by the fungal hydrophobin EAS. **109** (14), 2–9.
30. Kwan, A.H.Y., et al. (2006) Structural basis for rodlet assembly in fungal hydrophobins. *Proc. Natl. Acad. Sci.*, **103** (10), 3621–3626.
31. Bremer, P.J., et al. (2006) Laboratory scale Clean-In-Place (CIP) studies on the effectiveness of different caustic and acid wash steps on the removal of dairy biofilms. **106**, 254–262.
32. Simo, M., et al. (2005) Effect of mechanical stress on biofilms challenged by different chemicals. **39**, 5142–5152.
33. Zou, M., and Liu, D. (2018) A systematic characterization of the distribution , bio fi lm-forming potential and the resistance of the bio fi lms to the CIP processes of the bacteria in a milk powder processing factory. *Food Res. Int.*, **113** (January), 316–326.
34. Antoniou, K., and Frank, J.F. (2005) Removal of *Pseudomonas putida* Biofilm and Associated Extracellular Polymeric Substances from Stainless Steel by Alkali Cleaning. **68** (2), 277–281.
35. Jorgensen, H., et al. (1995) Metabolic Flux Distributions in *Penicillium chrysogenum* During Fed-Batch Cultivations. *Biotechnol. Bioeng.*, **46**, 117–131.

Chapitre 3

**Décryptage de la voie de biosynthèse de
l'astine par analyse bioinformatique**

1. Introduction

La production d'astine a été pendant très longtemps attribuée à la plante *A. tataricus* [1,2]. Différentes formes d'astine ont été détectées chez cette plante sans que la voie de biosynthèse correspondante ne soit mise en évidence [3]. L'identification des gènes impliqués dans la voie de biosynthèse de l'astine est une étape indispensable pour envisager d'accroître la productivité par ingénierie génétique, soit directement chez son producteur homologue, soit en transférant la voie chez un hôte hétérologue.

En 2015, une production d'astine a été mise en évidence chez *Cyanoderma asteris*, un champignon endophyte de la plante, ce qui a ouvert de nouvelles perspectives concernant l'élucidation de la voie de biosynthèse de l'astine [4,5]. Le génome de ce champignon filamenteux a été entièrement séquencé via la technologie haut-débit Illumina. L'assemblage et l'annotation du génome ont ensuite été réalisés respectivement par le biais des logiciels GS De Novo Assembler Software (Roche) et GenDB 2.0. Le génome obtenu, d'une taille totale estimée à 28,42 Mb, comprends 10.309 séquences codant pour des protéines [6].

Le génome de *C. asteris* présente une homologie de séquence de 95 % avec *Talaromyces islandicus*, un champignon filamenteux produisant deux pentapeptides cycliques, la cyclochlorotine (Cct) et l'hydroxy-cyclochlorotine (HCct), dont les structures sont proches de celles des astines [7]. L'astine C et la Cct ne diffèrent qu'au niveau du premier monomère reconnu lors de l'assemblage du NRP avec une incorporation de L-2Abu dans l'astine C et de L-Ser dans la Cct. Chez *T. islandicus*, les gènes *cctN* (16.926 pb) et *pisl3812-04907* (16.874 pb) ont été identifiés comme étant impliqués dans la biosynthèse de Cct et HCct. Les NRPSs correspondantes s'organisent en 5 modules (un pour chaque monomère incorporé) et sont terminées par un domaine de condensation terminal impliqué dans la cyclisation finale du NRP [7] (**Figure 1**).

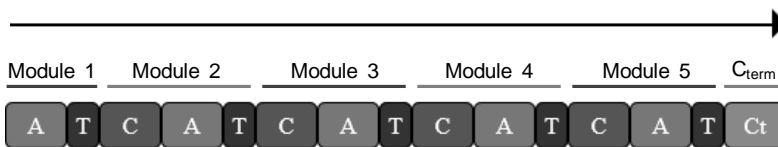


Figure 1 : Organisation modulaire des NRPSs, responsables de l'assemblage de la Cct et HCct, encodées par *cctN* et *pisl3812-04907* chez *T. islandicus*.

La très grande similarité de structure entre l'astine C et la Cct, ainsi que la forte homologie de séquence entre *C. asteris* et *T. islandicus*, laisse supposer que la synthétase d'astine présente une organisation très similaire à celle des synthétases de Cct et HCct identifiées chez *T. islandicus*. Cet élément constitue un point de départ vers l'élucidation de la voie de biosynthèse de l'astine chez *C. asteris*.

Ces dernières années, de nombreux outils bioinformatiques ont été développés afin de faciliter l'étude des voies de biosynthèse de type non-ribosomique [8]. Ces méthodes bioinformatiques ont été développées en se basant sur le fait qu'il existe une très forte conservation de séquence dans les domaines composant la synthétase. Les

séquences inconnues sont analysées par comparaison avec des séquences élucidées et répertoriées au sein des bases de données. L'utilisation combinée de différents outils bioinformatiques permet, à partir d'un génome séquencé, d'identifier les clusters de gènes relatifs à des voies NRPS, puis d'analyser l'organisation en modules / domaines des NRPSs, et enfin de prédire les spécificités de substrats des domaines identifiés [9].

Les outils développés permettent d'obtenir des analyses fiables concernant les NRPSs d'origine bactérienne, contrairement aux NRPSs fongiques où les résultats sont souvent imprécis. Ceci s'explique par un déséquilibre au sein des bases de données dans lesquelles les structures d'origine bactérienne sont majoritairement représentées. Par conséquent, les outils bioinformatiques utilisés à l'heure actuelle ont été développés en se fondant principalement sur des modèles bactériens, sans prendre en compte les spécificités des systèmes fongiques [10,11].

La synthétase d'astine, produite par un organisme fongique, appartient à cette dernière catégorie. Afin d'élucider son mode de fonctionnement et d'identifier les gènes impliqués dans la voie de biosynthèse de l'astine, le génome de *C. asteris* a été analysé selon un workflow alternatif de celui généralement utilisé pour les organismes bactériens. Ce chapitre détaille la mise en place de ce workflow, et les différents résultats obtenus au cours de cette analyse bioinformatique (**Figure 2**).

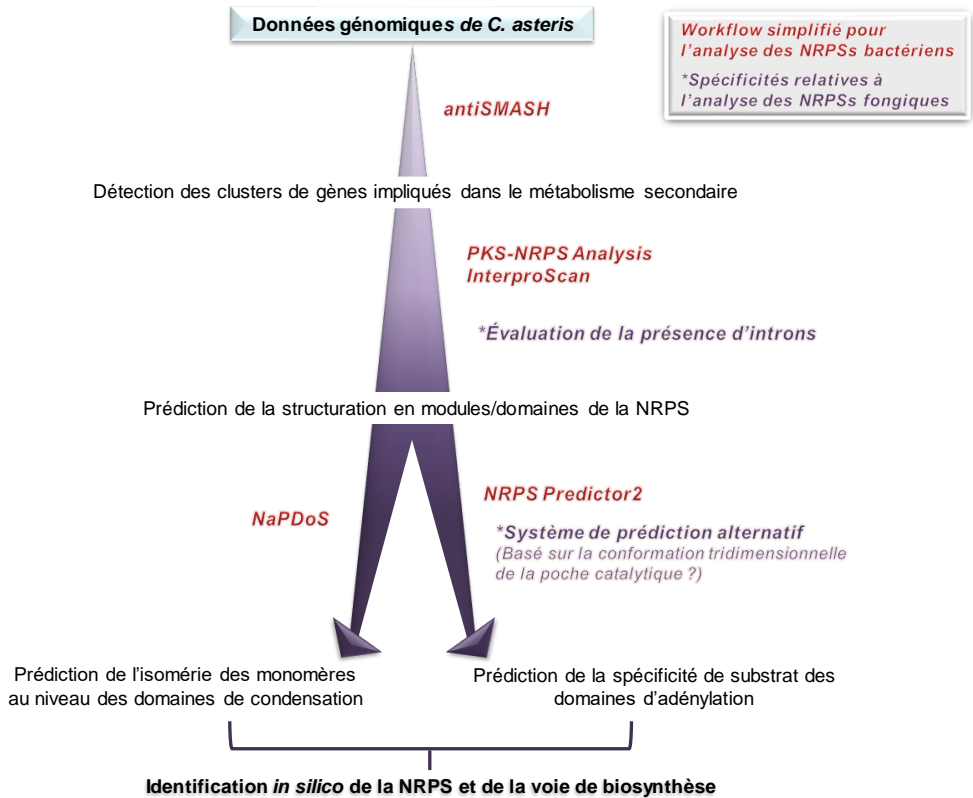


Figure 2 : Différentes étapes d'analyses bioinformatiques (tenant compte des spécificités relatives aux séquences fongiques) pour l'identification *in silico*, à partir des données génomiques de *C. asteris*, de la NRPS d'astine.

2. Matériels et méthodes

2.1. Analyse *in silico* des données génomiques

2.1.1. Analyse globale du génome

Le logiciel antiSMASH (<http://antismash.secondarymetabolites.org/>), version 4.0 [12], a été utilisé pour analyser le génome de *C. asteris*. Cet outil permet de rechercher à partir des données génomiques, l'ensemble des clusters de gènes impliqués dans la biosynthèse de métabolites secondaires (Terpènes, PKS, NRPS, PKS-NRPS, ou autres). Les données génomiques à analyser peuvent être téléchargées en tant que génomes entiers ou directement sous la forme d'ensemble de 'contigs' résultant du séquençage. Le logiciel propose une vue globale de l'ensemble des clusters identifiés. Pour chacun de ces clusters une page d'information est disponible détaillant les différents gènes présents dans le cluster et leur fonction supposée, la structuration en domaines enzymatiques de la NRPS / PKS, et une prédiction du métabolite secondaire potentiellement produit.

2.1.2. Analyse des différents gènes présents dans le cluster NRPS

En complément d'antiSMASH, l'outil BLASTp (« protein - protein Basic Local Alignment Search Tool », <https://blast.ncbi.nlm.nih.gov/Blast.cgi?Page=Proteins>), a été utilisé pour déterminer les fonctions supposées des protéines codées par les différents gènes des clusters NRPS. Ce logiciel a également été utilisé pour déterminer les pourcentages d'homologie de séquences entre les protéines codées par le cluster de *C. asteris* et celles codées par le cluster de la Cct chez *T. islandicus*. Cet outil réalise des alignements de séquences peptidiques afin de mettre en évidence des homologies entre la séquence fournie et les séquences présentes dans la base de données « non-redundant protein sequences » (nr) de NCBI (National Center for Biotechnology Information) ou celles fournies par l'utilisateur.

Selon le même principe et afin de compléter la première analyse, les séquences nucléotidiques des gènes présents dans les clusters NRPS ont été analysées via BLASTn (« nucleotide Basic Local Alignment Search Tool »). L'analyse des différents gènes a été réalisée avec le programme « discontinuous megablast », en considérant la base de données « nucleotide collection » (nr/nt).

Les séquences nucléotidiques et protéiques analysées pour chaque gène ont été identifiées et délimitées lors de l'analyse antiSMASH du génome de *C. asteris*.

2.2. Prédiction de la structuration en modules / domaines des NRPSs

Les prédictions relatives à la structuration en modules / domaines ont été réalisées en combinant les résultats obtenus avec antiSMASH, PKS/NRPS Analysis (<http://nrps.igs.umaryland.edu/>), et InterProScan (<http://www.ebi.ac.uk/interpro/search/sequence-search>). PKS/NRPS Analysis et InterProScan détectent des signatures spécifiques au sein de séquences protéiques afin d'attribuer une activité enzymatique. InterProScan est un outil permettant de réaliser une recherche de domaines pour tout type de protéines, tandis que PKS/NRPS Analysis est dédié uniquement à l'analyse des protéines modulaires (PKS et NRPS). La conversion des séquences nucléotidiques des NRPSs en séquences protéiques (format accepté par PKS/NRPS Analysis et InterProScan) a été réalisée via la fonction « translate » disponible sur le portail ExPaSy, regroupant plusieurs outils bioinformatiques (<https://web.expasy.org/translate/>).

2.3. Prédiction des peptides formés par analyse des domaines enzymatiques des NRPSs

2.3.1. Prédiction de l'isomérisation des monomères

La prédiction de l'isomérisation des monomères incorporés au sein du NRP a été réalisée en se basant sur une analyse des propriétés des domaines de condensation composant la NRPS. La séquence protéique de chaque domaine de condensation (délimitée par antiSMASH), en format FASTA, a été indépendamment analysée via le logiciel NaPDoS (http://npdomainseeker.sdsc.edu/run_analysis.html). Cet outil permet de

déterminer à quel sous-type appartient le domaine C analysé (${}^L\text{C}_L$, ${}^D\text{C}_L$, C_{start} , $\text{C}_{\text{hétérocyclisation}}$, Dual C/E, ou E). Le domaine d'épimérisation E peut être considéré d'un point de vue structural comme un sous-domaine de condensation. L'analyse est basée sur un système de comparaison avec des séquences et signatures préalablement identifiées comme étant spécifiques de l'un ou l'autre des sous-types de domaine C.

2.3.2. Analyse de la spécificité des domaines d'adénylation

Certains outils bioinformatiques cités précédemment tels qu'antiSMASH et PKS/NRPS Analysis fournissent des éléments de prédiction quant à la nature du monomère activé par les domaines d'adénylation. Ces résultats ont été complétés en analysant indépendamment la séquence protéique des domaines d'adénylation via le logiciel NRPS Predictor 2 (<http://nrps.informatik.uni-tuebingen.de/>) [13]. Cet outil est spécifiquement dédié à la prédiction de la spécificité des domaines d'adénylation. Les prédictions fournies sont classifiées selon leur niveau de précision allant d'une simple estimation des propriétés physico-chimiques du monomère reconnu, jusqu'à la prédiction d'un substrat unique (« three clusters », « large clusters », « small clusters », et « nearest neighbor »).

2.4. Modélisation tridimensionnelle de la NRPS

2.4.1. Établissement d'un modèle 3D de la NRPS

Les matrices de références, souvent anglicisées par « templates », nécessaires à l'établissement d'un modèle, ont été identifiées grâce aux modules de recherche de templates des trois logiciels HHpred (<http://toolkit.tuebingen.mpg.de/#/tools/hhpred>) [14], Phyre2 [15] et I-TASSER [16] dans le but de maximiser l'identification d'homologues. Ces logiciels permettent de réaliser des comparaisons entre une séquence protéique en entrée et la base de données Protein Data Bank (PDB) grâce à différents algorithmes basés notamment sur des profils protéiques. Les meilleurs templates en score et en couverture étaient : 2VSQ, 4ZXH et 5T3D. Les trois templates identifiés étaient très proches entre eux pour les trois logiciels utilisés.

Les logiciels Phyre 2 en mode intensif et I-TASSER ont ensuite été utilisés avec les paramètres par défaut pour générer par homologie des modèles tridimensionnels de chaque module de la synthétase, en se basant sur les templates précédemment identifiés. Ces deux logiciels présentent la particularité de modéliser *ab initio* toute portion de protéine qui ne présenterait aucun template, contrairement à HHpred, par exemple. Phyre2 génère un seul modèle final, alors qu'I-TASSER peut en générer jusqu'à 5. Les modèles de Phyre2 et I-TASSER ont été visualisés et alignés grâce au logiciel PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) dans le but d'évaluer d'éventuelles divergences. La qualité des modèles ainsi générés a également été soumise à Procheck [17]. Les modèles étant globalement très proches les uns des autres lors d'une inspection visuelle, mais ceux de Phyre2 présentant des angles ϕ et ψ optimisés dans les graphes de Ramachandran, seuls les modèles de Phyre2 de chaque module ont été conservés.

2.4.2. Évaluation de la spécificité des domaines d'adénylation sur base du modèle tridimensionnel de la NRPS

Les modèles des domaines d'adénylation des différents modules de l'astine synthétase ont été générés grâce à Swissprot avec le template 1AMU. Ce logiciel permettait d'intégrer l'AMP dans les modèles avec un minimum de « clashes ». La position de l'AMP au sein des différents modèles a été réalisée par le biais d'une superposition avec 1AMU où est intégré une molécule d'AMP. Ces modèles ont ensuite été utilisés pour étudier l'amarrage moléculaire, communément appelé « docking », entre les substrats et les domaines d'adénylation. Un score de couplage enzyme-substrat a été calculé entre les modules modélisés et les monomères réellement incorporés par la NRPS dans le peptide d'astine C (L-2Abu, L-Pro, L-2Abu, L-Ser, β -Phe). À titre de comparaison, la même analyse a été réalisée avec les monomères prédits par le logiciel NRPS Predictor2 (Val, Leu, Gln, Cys, Asp). Les scores de « docking » ont été déterminés pour chaque ligand en présence de l'AMP, qui se lie au monomère lors de la réaction d'adénylation. Les analyses ont été réalisées à l'aide du logiciel GOLD [18], adapté pour réaliser un « docking » entre des protéines et des ligands de faible poids moléculaire. Le docking a été réalisé en considérant une sphère de 15 Ångströms centrée sur le phosphate du cofacteur (AMP), avec le champ de force ChemPLP, avec la protéine paramétrée comme rigide et les ligands flexibles. Les ligands soumis au docking ont été classés par score décroissant, et le ligand de meilleur score a été considéré comme le ligand le plus favorable pour la réaction d'adénylation dans le module.

3. Résultats

3.1. Analyse des données génomiques de *C. asteris*

3.1.1. Recherche de clusters de gènes impliqués dans la synthèse de métabolites secondaires


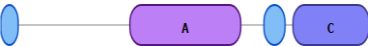
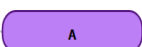
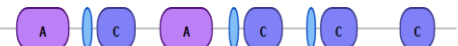

L'analyse du génome de *C. asteris* via antiSMASH a permis de détecter 38 clusters de gènes impliqués dans la biosynthèse de métabolites secondaires. Parmi les clusters détectés, 5 ont été associés à la voie non-ribosomique et peuvent donc être considérés comme étant potentiellement impliqués dans la production d'astine (**Tableau 1**).

Tableau 1 : Nombre de clusters détectés lors de l'analyse antiSMASH du génome de *C. asteris*, répartis selon le type de métabolites secondaires associés à ces clusters.

Type de cluster de gènes	Nombre de clusters détectés
NRPS	5
PKS	13
NRPS-PKS	1
Terpènes	6
Autres	13

AntiSMASH propose pour chacune des séquences NRPS une prédiction de la structuration en domaines. Les résultats obtenus pour chacun des clusters NRPS sont présentés sur le **Tableau 2**.

Tableau 2 : Annotation détaillée des domaines composant les NRPSs identifiées dans les 5 clusters détectés par antiSMASH dans le génome de *C. asteris*.

N° de cluster	Position du cluster dans le génome	Annotation détaillée des domaines	Nombre de modules
6	ctg45_orf8		5
7	ctg46_orf21		1
9	ctg49_orf10		1
15	ctg105_orf9		4
38	ctg488_orf0002		2

Les NRPSs encodées par les clusters n° 7, 9, et 38 présentent une structuration avec un ou deux modules, et sont donc incapables d'assembler de manière linéaire un peptide de 5 monomères tel que l'astine. En ce qui concerne la NRPS du cluster n° 15, une structuration avec 4 modules a été identifiée. Une analyse complémentaire réalisée via PKS/NRPS Analysis a confirmé les résultats obtenus via antiSMASH. Seuls 2 domaines d'adénylation capables d'activer uniquement deux substrats différents sont présents au sein de cette synthétase. Celle-ci ne peut donc pas être associée à la biosynthèse de l'astine C qui comporte 5 monomères (L-2Abu, L-Pro, L-2Abu, L-Ser, β -Phe).

Le cluster n° 6 (ctg45_orf8) code pour une NRPS présentant une structure en adéquation avec ce qui est attendu pour la synthétase d'astine. L'organisation de cette synthétase est similaire à celle observée pour les NRPSs de Cct et HCct chez *T. islandicus*. Si cette synthétase fonctionne selon un mode d'assemblage linéaire, les 5 domaines d'adénylation présents sont potentiellement impliqués dans l'assemblage d'un pentapeptide. Cette NRPS est terminée par un domaine de condensation terminal dont l'implication dans le processus de cyclisation du NRP produit a été démontrée chez d'autres systèmes NRPS fongiques [7,19]. Des analyses complémentaires ont été réalisées sur ce cluster n° 6 afin de confirmer son implication dans la biosynthèse d'astine.

3.1.2. Organisation du cluster n° 6 (ctg45_orf8) de *C. asteris*

Le cluster n° 6 comporte un total de 11 gènes dont le gène codant pour la synthétase (*astS*). La disposition des différents gènes au sein du cluster ainsi que les enzymes associées et leur fonction supposée sont présentées **Figure 3**.

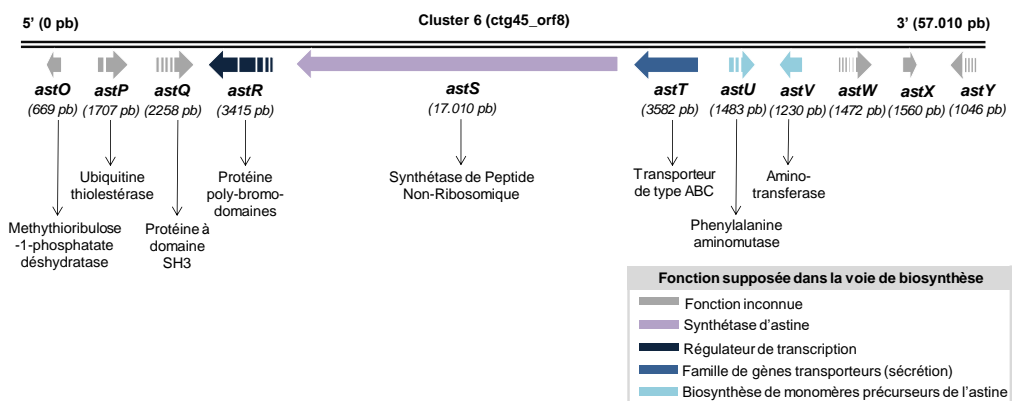


Figure 3 : Organisation du cluster n° 6 (ctg45_orf8) de *C. asteris* contenant le gène *astS* correspondant potentiellement à la synthétase d'astine. Les rôles supposés, dans la biosynthèse de l'astine, des protéines encodées par les différents gènes du cluster ont été évalués via les outils BLASTp et BLASTn.

En utilisant l'outil BLASTp, les séquences des différents gènes du cluster ont été alignées avec les séquences répertoriées dans la base de données « nr » de NCBI.

Lorsque la séquence analysée présente un pourcentage d'homologie élevé avec une protéine de la base de données dont l'activité a été caractérisée, il est possible par analogie de déduire l'activité potentielle de l'enzyme codée. Certains gènes tels que *astO*, *astW*, *astX*, et *astY* ne présentent pas d'homologie de séquence avec des structures déjà répertoriées. Par conséquent leur rôle dans la biosynthèse du NRP reste indéterminé. L'analyse complémentaire des séquences nucléotidiques via BLASTn a montré qu'un fragment de la séquence *astO* (~80 %) présente une homologie de 67 à 69 % avec des gènes codant pour une méthylthioribulose-1-phosphate déshydratase, identifiés chez *Paraphaeosphaeria sporulosa*, *Diploidia corticola*, et *Alternaria arborescens*. Cette enzyme est impliquée dans la voie de régénération de la méthionine où elle catalyse la déshydratation de la méthylthioribulose-1-phosphate [20]. L'implication d'une telle enzyme dans la synthèse d'astine n'a pas été démontrée.

La protéine codée par *astQ* comporte un domaine SH3 qui est un motif très répandu chez les protéines-signal d'origine eucaryote [21]. La présence de SH3 est notamment caractéristique des protéines membranaires, ou impliquées dans la régulation des interactions protéine-protéine à proximité des membranes plasmiques [22]. La fonction de ce domaine n'ayant pas été déterminée, aucune information ne permet d'évaluer son rôle, ainsi que celui de la protéine où il s'inscrit, au sein de la voie de biosynthèse de l'astine.

Un fragment de la protéine codée par le gène *astP* (~84 %) présente 54 % d'homologie avec une ubiquitine thiolestérase produite par la levure noire *Exophiala dermatitidis*. L'ubiquitine est une petite protéine (8 kDa) capable de se conjuguer aux résidus lysines des protéines. Cette ubiquitination constitue un signal induisant, entre autres, la dégradation de la protéine complexée par le protéasome [23]. L'ubiquitine thiolestérase appartient à la famille des enzymes de déubiquitination, capable de désactiver l'ubiquitine en hydrolysant la partie thiol-ester, préservant ainsi la protéine de la protéolyse [24]. L'enzyme codée par *astP* est donc potentiellement responsable de la protection, vis-à-vis du système protéolytique de *C. asteris*, de la NRPS AstS et des autres enzymes impliquées dans la voie de biosynthèse.

La séquence protéique de AstR comporte des éléments identifiés comme codant pour des bromo-domaines. Ces structures sont spécifiques de séquences impliquées dans la régulation de la transcription [25,26]. Le gène *astT* présente une homologie de séquence avec des gènes codant pour des transporteurs membranaires de type ABC (« ATP binding cassette »). Il est probable que ces transporteurs ABC soient impliqués dans le transport et la sécrétion de l'astine chez *C. asteris*.

3.1.3. Comparaison entre le cluster n° 6 de *C. asteris* et le cluster de gènes associé à la synthèse de la cyclochlorotine chez *T. islandicus*

La cyclochlorotine synthétisée par *T. islandicus* est un pentapeptide cyclique similaire à celui de l'astine C où seul le monomère reconnu par le premier domaine d'adénylation diffère (L-Ser pour la Cct et L-2Abu pour l'astine C). Les deux NRPs étant très proches, il est intéressant de comparer le cluster de gènes associé à la synthèse de cyclochlorotine avec celui identifié chez *C. asteris* comme étant potentiellement impliqué dans la synthèse de l'astine [7] (**Figure 4**).

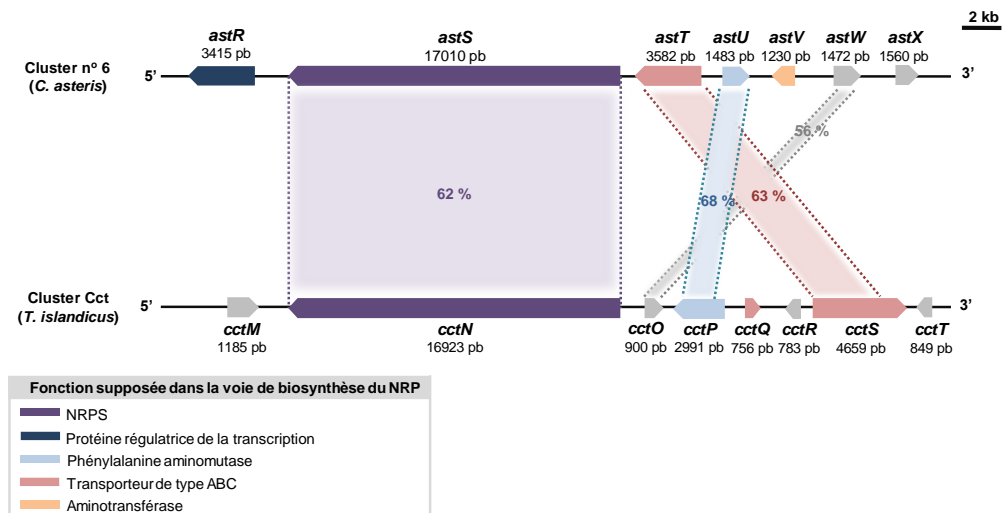


Figure 4 : Analyse comparative des gènes présents au sein du cluster n° 6 (contenant le gène *astS* codant potentiellement pour la NRPS de l'astine) chez *C. asteris* et ceux identifiés dans le cluster de gènes responsable de la synthèse de la cyclochlorotine *T. islandicus*. Le pourcentage d'homologie entre les protéines encodées par certains gènes des deux clusters sont indiquées.

L'organisation du cluster n° 6 chez *C. asteris* présente quelques similitudes avec celle précédemment décrite chez *T. islandicus* pour le cluster de gènes impliqué dans la biosynthèse de la cyclochlorotine [7]. Au sein des deux clusters, un large gène codant pour une NRPS est présent. Les synthétases codées par ces gènes présentent 62 % d'homologie au niveau de leur séquence d'acides aminés. Les protéines codées par *astW* et *cctO* dont la fonction au sein de la voie de biosynthèse de l'astine et de la Cct n'a pas été identifiée, présentent 56 % d'homologie. Par ailleurs, les protéines codées par *astT* et *cctS*, prédites en tant que transporteur de type ABC, présentent 63 % d'homologie. Une phénylalanine aminomutase, potentiellement impliquée dans la synthèse du monomère de β -Phe, est codée par le gène *astU* chez *C. asteris*, et *cctP* chez *T. islandicus*. Les deux protéines présentent 68 % d'homologie. En revanche, l'aminotransférase identifiée au sein du cluster n° 6 chez *C. asteris* et potentiellement impliquée dans la formation du L-2Abu, ne semble pas être codée au sein du cluster de la Cct.

Les similarités observées entre les deux clusters, ainsi que les homologies de séquences de plusieurs protéines codées par le cluster n° 6 avec des protéines impliquées dans la synthèse de la Cct, conforte la très probable implication du cluster n° 6 dans la biosynthèse d'un NRP comme l'astine.

3.2. Analyse des séquences nucléotidique et protéique de *AstS*

3.2.1. Prédiction de la structuration en modules / domaines

Le logiciel antiSMASH, dont les résultats ont été présentés ci-dessus, propose une organisation de la synthétase encodée par *astS* (Cluster n° 6 - **Tableau 2**). Des analyses complémentaires ont été réalisées afin de confirmer la structuration de cette NRPS. Les résultats obtenus à l'aide des logiciels InterProScan et PKS/NRPS Analysis sont présentés **Figure 5**.

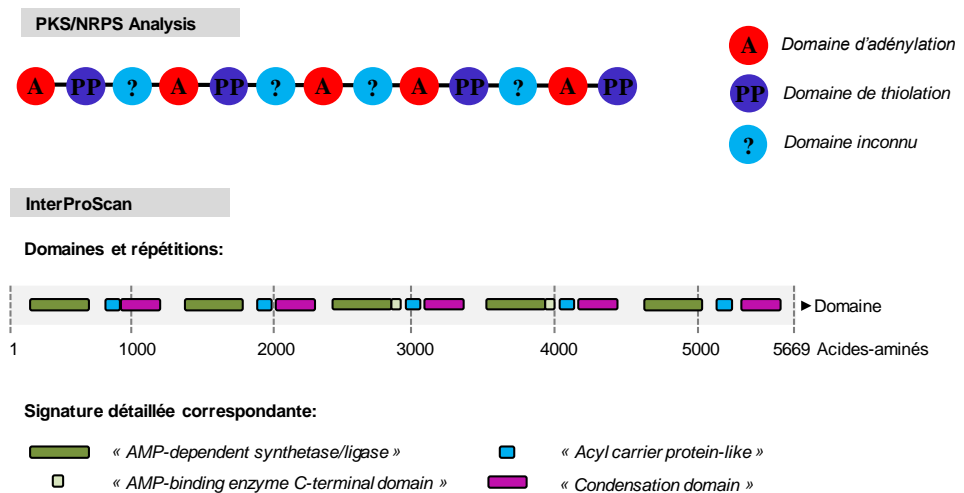


Figure 5 : Analyse de la séquence protéique de la NRPS encodée par *AstS* via PKS/NRPS Analysis et InterProScan.

La détection des domaines par PKS/NRPS Analysis est moins complète que celle obtenue via antiSMASH. En particulier, le domaine T du module 3 et le domaine de condensation terminal n'ont pas été détectés. Les autres domaines de condensation ont bien été détectés mais n'ont pas été reconnus comme tels et ont été annotés en tant que « domaine inconnu ».

L'analyse InterProScan de la NRPS *AstS* a fourni des résultats similaires à ceux obtenus en utilisant antiSMASH. Les quatre domaines de condensation, ainsi que le domaine de condensation terminal ont correctement été détectés. Les 5 domaines de thiolation, aussi appelés « Peptidyl Carrier Protein » (PCP) ou « Acyl Carrier Protein » (ACP) en référence à leur fonction, ont été identifiés et annotés en tant qu'ACP. De même, cinq domaines « AMP-dependent synthetase » ont été détectés.

La fonction « AMP-dependent » est caractéristique des domaines d'adénylation, qui catalysent la fixation d'un groupement AMP sur les monomères reconnus.

En compilant les résultats obtenus à partir d'antiSMASH, InterProScan, et PKS/NRPS Analysis, la structuration de la synthétase codée par *astS* a été prédite comme suit : A-T-C-A-T-C-A-T-C-A-T-C-A-T-C_{term}. Cette organisation est similaire à celle décrite pour les NRPSs de Cct et HCct identifiées chez *Talaromyces islandicus* (**Figure 1**).

3.2.2. Impact de la présence d'introns sur l'activité prédite des domaines

L'annotation du gène *astS* à la suite du séquençage du génome de *C. asteris* a mis en évidence l'existence de deux introns de 45 pb chacun. Ces deux introns sont localisés au sein des domaines d'adénylation. Le premier est présent dans le module 1, et le second dans le module 3. Les deux introns étant des multiples de 3, le cadre de lecture ne sera pas affecté par l'épissage, mais celui-ci générera des domaines d'adénylation avec une séquence protéique raccourcie de 15 acides aminés.

Deux séquences protéiques sont donc possibles concernant les domaines d'adénylation 1 et 3, selon que les séquences identifiées comme « introns » soient épissées ou non lors des modifications post-transcriptionnelles. Si les 45 nucléotides sont épissés lors de la maturation de l'ARNm, l'intron est réel et la protéine générée sera raccourcie de 15 acides aminés. Au contraire, si l'intron résulte d'une erreur d'annotation, la séquence ne sera pas excisée lors de l'épissage et la protéine conservera ces 15 acides aminés.

Les domaines d'adénylation sont caractérisés par 10 signatures (numérotées de A1 à A10) généralement détectables au sein de la séquence protéique (**Tableau 1 - Chapitre 1**). Une analyse des signatures présentes au sein de la séquence protéique avec ou sans épissage des introns (et perte des 15 acides aminés correspondants) a été réalisée pour les deux domaines d'adénylation concernés. Les signatures identifiées dans chaque cas sont les séquences mises en évidence par les couleurs rouge et verte sur la **Figure 6**.

Module 1 : Domaine A sans épissage					vs	avec épissage				
LATDLPEQRL	DQIAAVIQPN	CALALASTGQ	EGRATHLSKT	VLEVDSRLSE	50	LATDLPEQRL	DQIAAVIQPN	CALALASTGQ	EGRATHLSKT	VLEVDSRLSE
ESDISVASAD	SDSHVGPPEGL	ICLIFTSGTT	GLPKAAMLTH	AGYCALAANI	100	ESDISVASAD	SDSHVGPPEGL	ICLIFTSGTT	GAANGIPLD	IGPGAKVLNW
GPLEDIGPGA	KVLNWAAAYAW	DVSIADNMVS	LLTGACLTIP	STWETKSNLT	150	AAAYAWDVSLA	DNMVSLLTGA	CLTIPSTWET	KSNLTETIAH	YQATHVLITP
ETIAHYQATH	VLITPQVSKL	IVPSEVPSLK	VVALGGPECT	KDATERWRGT	200	QVSKLIVPSE	VPSLKVVVALG	GEPCTKDATE	RWRGTGVKLM	NVYGPACESC
GVKLMNVYGP	AESCSCVTVAH	QDVVNAPTSP	SIGRGFSATW	IVDPDNVDRL	250	VTVAHQDVVN	APTSPSIRG	FSATWIVDPD	NVDRLLPVGA	IGELVLEGPQ
LPVGAIGELV	LEGQVQPGY	LHDAKNTAAA	YLDNPQWLVE	GRQGVAPGRS	300	VGPGYLHDAK	NTAAAYLDNP	QWLVEGRQGV	APGRSGRLYR	TGDLVKYQDI
GRLYRTGDLV	KYQDDGQIDI	CGRRDHQVQV	RQRIELGEL	GTHLQRLIPE	350	GQIDICGRRD	HQVKVRGQRI	ELGELGTHLQ	RLIPEEKWS	TDVAKLANGS
EIKWSTDVAK	LANGSEMLLV	FVVLPRRIES	SFADDTLADT	IRDANTDLAA	400	EMLLVFPVLP	PRIESSFADD	TLADTIRDAN	TDLAARLPSA	LVPGAYVKLF
RLPSALVPGA	YVKLDEIPLS	PNGKTRRRLK	REIAASLSTE	QLMM		EIPLSPNGKT	DRRKLREIAA	SLSTEQLMM		

Module 3 : Domaine A sans épissage					vs	avec épissage				
MDVDYPEQRL	KTICELMKAR	IILASPAQQE	LAKRLVNVF	VVERSCTTL	50	MDVDYPEQRL	KTICELMKAR	IILASPAQQE	LAKRLVNVF	VVERSCTTL
GTASLDVVDS	TDTKRMYIC	FTSGSTGVPK	GVVVAHRNLA	SASVQQVRDL	100	GTASLDVVDS	TDTKRMYIC	FTSGSTGVPK	GVVVAHRNLA	SASVQQVRDL
GFTPEDRVYD	FSSHFDACI	WHWYGLVAG	ATVCTPSQES	RMNLAGDIT	150	GFTPEDRVYD	FSSHFDACI	WHWYGLVAG	ATVCTPSQES	RMNLAGDIT
SFNTTKLFLT	PSVARTINPR	EVPTVKRMWL	GGEAVTPLDV	SMWKDHLDLW	200	SFNTTKLFLT	PSVARTINPR	EVPTVKRMWL	GGEAVTPLDV	SMWKDHLDLV
CAYGPTIETP	LCMFTPLSRP	EAASNIGKGS	GVLPWVCNPS	NYNELVALGA	250	CAYGPTIETP	LCMFTPLSRP	EAASNIGKGS	GVLPWVCNPS	NYNELVALGA
VGEMVIEGPL	VTQGYHQQPE	KTAQVFIEDP	EFLRGSKSR	PGRHGRLYRT	300	VGEMVIEGPL	VTQGYHQQPE	KTAQVFIEDP	EFLRGSKSR	PGANGSIQYL
GDLVRYTANG	SIQYLGRGDT	QVKLRGQRVE	FGEIEYQLKH	ALPESPAVVC	350	GRGDTQVKLR	QQRVEFGEIE	YQLKHALPES	PAVVECVIHI	PTGRPALVAF
EVVIHTGRP	ALVAFCANPS	GSNLDIPRAR	MHLKTKLPPY	MI		CANPSGNSLD	IPRARMHLKT	KLPPYMI		

Figure 6 : Séquences protéiques des domaines d'adénylation des modules 1 et 3 obtenues après épissage ou non des séquences identifiées comme introns. Les signatures caractéristiques sont mises en évidence par la couleur rouge. Les signatures caractéristiques localisées dans les zones annotées comme introns sont présentées en vert.

L'épissage de l'intron au niveau du domaine d'adénylation du module 1 entraîne la disparition de l'enchaînement d'acides aminés « L-I-C-L-I-F-T-S-G-T-T-G-T-P-K-A-A » (indiqué en vert). Cette séquence présente un taux de similarité de 75 % avec la signature A3 « L-A-Y-x-x-Y-T-S-G-[ST]-T-G-x-P-K-G ». En ce qui concerne le domaine d'adénylation du module 3, la séquence d'acides aminés « Y-R-T-G-D-L » n'est plus présente en cas d'épissage de l'intron. Le taux de similarité de celle-ci avec la signature A7 « Y-[RK]-T-G-D-L » est de 100 %.

Dans les deux cas, l'épissage des séquences initialement annotées comme introns entraîne la disparition d'une des signatures caractéristiques de la chaîne peptidique des domaines d'adénylation. Ces signatures étant des motifs très conservés d'un domaine A à un autre et d'un organisme à un autre, on peut supposer que leur présence est nécessaire pour la structuration des domaines d'adénylation [27]. Les introns identifiés lors de l'annotation du génome sont donc probablement de faux positifs. L'absence d'épissage de ces séquences par *C. asteris* a été confirmée expérimentalement par la suite, par l'un des partenaires du projet ERA-IB-15-039 (ASTINPROD), en séquençant l'ADNc obtenu par transcription inverse de l'ARNm de *C. asteris* (données non incluses).

3.3. Prédiction de la spécificité des domaines de la NRPS AstS

3.3.1. Détermination du sous-type des domaines de condensation

Les domaines de condensation ont été analysés via le logiciel NaPDoS afin de déterminer leur sous-type (**Tableau 3**).

Tableau 3 : Résultat de l'analyse réalisée avec NaPDoS pour les 5 domaines de condensation identifiés au sein de la NRPS AstS, et les 2 domaines C_{term} des NRPSs de Cct et HCct (T.isl1 et T.isl2).

ID de Requête	ID correspondance de la base de données	Identité (%)	Longueur d'alignement	e-valeur	Voie de biosynthèse	Type de sous-domaine
AstS.C1	hctox1_C3_dual	28	272	1e-25	HC-toxin	dual
AstS.C2	hctox1_C3_dual	29	274	9e-29	HC-toxin	dual
AstS.C3	hctox1_C3_dual	27	251	6e-23	HC-toxin	dual
AstS.C4	hctox1_C3_dual	27	244	7e-18	HC-toxin	dual
AstS.C _{term}	cyclo1_C12_dual	25	254	1e-22	Cyclosporin	dual
T.isl1-C _{term}	cyclo1_C12_dual	22	355	1e-21	Cyclosporin	dual
T.isl2-C _{term}	cyclo1_C12_dual	24	360	3e-27	Cyclosporin	dual

Les cinq domaines de condensation de l'AstS ont été identifiés en tant que domaine « dual C/E », capable de catalyser la condensation entre deux monomères et de modifier l'isomérisation de l'un des monomères (de la forme L à la forme D). Cette classification est surprenante dans la mesure où aucun acide aminé sous la forme D n'est incorporé dans le peptide d'astine. Les 4 premiers domaines de condensation établissent une liaison peptidique entre des L-monomères activés par des modules successifs. Par conséquent, la classification attendue pour ces 4 domaines serait plutôt de type ^LC_L (établissant la liaison peptidique entre deux monomères sous la forme L). En revanche, le domaine de condensation terminal, également identifié en tant que « dual C/E », ne peut pas catalyser la réaction de condensation entre deux monomères issus de modules successifs. Celui-ci est situé à l'extrémité de la NRPS et n'est donc pas suivi d'un domaine A activant un monomère.

Pour chaque domaine de condensation analysé, NaPDoS établit une correspondance avec le domaine le plus proche au sein de sa base de données (en termes de séquences protéiques). Les résultats obtenus pour le domaine C_{term} et les 4 autres domaines de condensation divergent. Les 4 premiers domaines présentent entre 27 et 29 % d'identité avec le domaine de condensation 3 de la synthétase de la toxine HC. Cette toxine, produite par le champignon *Cochliobolus carbonum*, est un NRP cyclique composé de 4 monomères (D-Pro, L-Ala, D-Ala, et L-Aeo) [28]. Le domaine de condensation 3, catalysant la liaison entre les monomères D-Ala et L-Aeo, a été identifié en tant que « dual C/E ». Ce domaine catalyse donc à la fois l'épimérisation

de L-Ala en D-Ala et l'établissement de la liaison peptidique avec L-Aeo. Les 4 premiers domaines de condensation de la NRPS AstS sont donc associés à un sous-type « dual C/E », qui ne correspond pas à la réalité biologique dans le cas de l'assemblage de l'astine C. Néanmoins, le faible pourcentage d'identité associé (de 27 à 29 %) souligne le manque de robustesse des prédictions réalisées.

Les résultats concernant le domaine C_{term} sont différents puisque celui-ci est associé au domaine de condensation 12 de la synthétase de la cyclosporine avec un score d'identité de 25 %. La présence d'un domaine de condensation terminal chez les NRPSs d'origine fongique est très souvent associée à un processus de cyclisation du peptide formé [19]. Les résultats obtenus lors de l'analyse NaPDoS pour ce domaine C_{term} de l'AstS ont été comparés avec ceux obtenus pour les domaines C_{term} des NRPSs de Cct et HCct (nommés T.isl1 et T.isl2) (**Tableau 3**). Les résultats obtenus sont similaires pour les 3 domaines C_{term} considérés. Un arbre phylogénétique, basé sur les séquences répertoriées dans la base de données NaPDoS, a été généré afin de confirmer les différences structurales entre les domaines C_{term} et les autres domaines C de l'AstS (**Figure 7**).

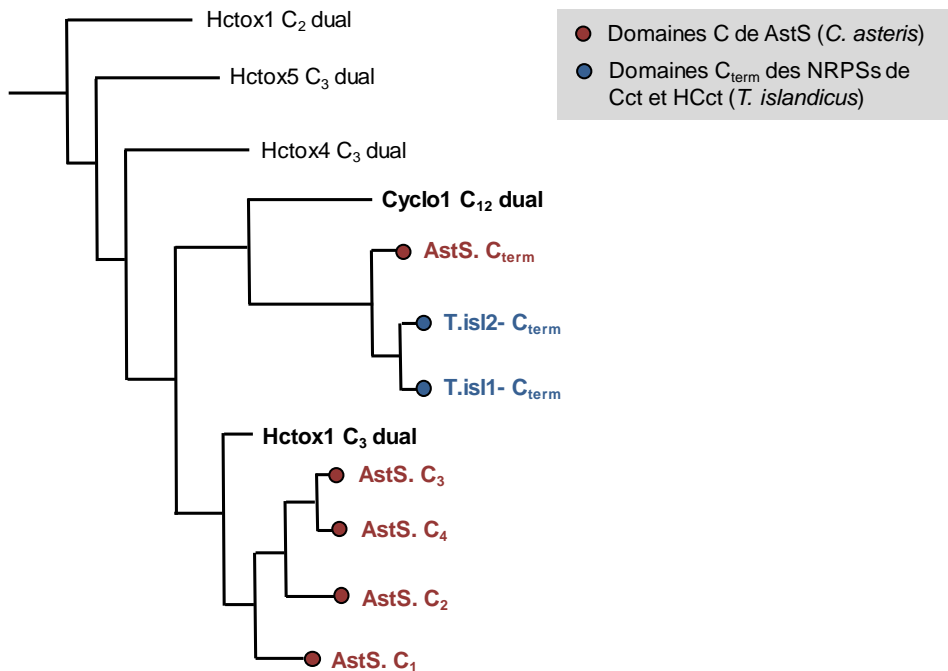


Figure 7 : Portion de l'arbre phylogénétique généré via NaPDoS, mettant en évidence la conservation des séquences protéiques entre les domaines de condensation composant AstS, les domaines C_{term} T.isl1 et T.isl2, et les domaines répertoriés dans la base de données.

Les 3 domaines C_{term} se placent sur une branche distincte (phylogénétiquement proche de cyclo1_C12_dual) de la branche des autres domaines de condensation (associés à hctox1_C3_dual). Les C_{term} des deux NRPSs de *T. islandicus* présentent

un score de conservation élevé entre eux, mais également vis-à-vis du C_{term} de la NRPS AstS. Les séquences de ces trois C_{term} et celui de la cyclosporine synthétase ont été alignées pour représenter les zones d'homologie de séquences. Il existe une très grande similarité entre le C_{term} de l'AstS et les C_{term} T.isl1 et T.isl2. En particulier, entre le C_{term} de AstS et T.isl1, les séquences protéiques sont homologues à 62 % (sur les 247 acides aminés composant le C_{term} de AstS). À titre de comparaison, le premier domaine de condensation de l'AstS ne présente que 21 % d'homologie de séquences avec ce même domaine C_{term} T.isl1.

Une étude réalisée précédemment avait mis en évidence l'absence de signature spécifique des sous-types Dual C/E, C_{start} , ${}^L C_L$, et ${}^D C_L$ chez le domaine C_{term} des deux NRPSs de *T. islandicus* [29]. Étant donné la très forte similitude de ces deux domaines avec le C_{term} de l'AstS, il est très probable que ces 3 domaines appartiennent au même sous-type de domaine C, et soient impliqués dans la cyclisation de leurs NRPs respectifs. Néanmoins, l'absence de signatures spécifiques pour ce type de domaine empêche d'attribuer formellement ces domaines au type des C_{term} impliqué dans la cyclisation finale des NRPs fongiques.

3.3.2. Prédiction de la spécificité de substrat des domaines d'adénylation

3.3.2.1. Méthode de prédiction traditionnelle

En synthèse non-ribosomique, les différents monomères sont recrutés et activés par les domaines d'adénylation de la synthétase correspondante. La spécificité de substrat des domaines d'adénylation peut être prédite en analysant leur site de liaison au site actif. Le logiciel NRPS Predictor2 permet d'analyser les propriétés physico-chimiques de 34 résidus, dans un rayon de 8 Ångström autour du site actif, afin de prédire la spécificité des domaines A. Les séquences protéiques des domaines d'adénylation présentes dans la NRPS AstS ont été analysées via ce logiciel (**Tableau 4**).

Tableau 4: Prédictions de la spécificité des domaines d'adénylation de la NRPS AstS via NRPS Predictor2 (np : non prédit)

Domaine	Prédictions NRPS Predictor2				Monomère incorporé dans l'astine
	« Three clusters »	« Large clusters »	« Small clusters »	« Nearest neighbor »	
A1	hydrophobic-aliphatic	gly, ala, val, leu, ile, abu, iva	pro	Val (50 %)	L-2Abu
A2	hydrophobic-aliphatic	np	np	Leu (70 %)	L-Pro
A3	hydrophobic-aliphatic	gly, ala, val, leu, ile, abu, iva	val, leu, ile, abu, iva	Gln (50 %)	L-2Abu
A4	hydrophobic-aliphatic	np	np	Cys (40 %)	L-Ser
A5	hydrophobic-aliphatic	np	np	Asp (50 %)	β-Phe

Les monomères prédits à l'échelle « Nearest neighbor » ne correspondent pas à ceux incorporés dans l'astine C (L-2Abu, L-Pro, L-2Abu, L-Ser, β -Phe). Dans la catégorie « Large cluster », l'acide aminobutyrique (abu) est identifié comme l'un des monomères potentiellement reconnu et activé par les domaines d'adénylation 1 et 3. La prédiction relative au domaine d'adénylation 4 positionne la cystéine en tant que monomère le plus probable. La L-Cys est un acide aminé dont la conformation est très proche de la L-Ser, qui est incorporée dans l'astine C (**Figure 8**).

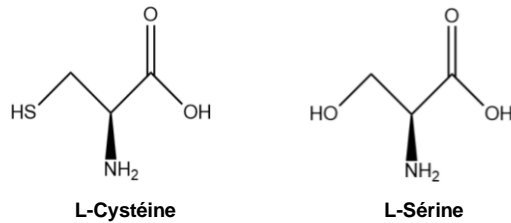


Figure 8 : Structures des deux acides aminés : L-cystéine et L-Sérine.

Les deux acides aminés ne varient qu'au niveau d'un seul groupement (sulfhydryle pour la L-Cystéine et hydroxyle pour la L-Sérine). En termes de prédiction de spécificité, la proximité structurale entre ces deux résidus peut compliquer le processus de discrimination.

Les prédictions obtenues concernant la spécificité des domaines d'adénylation 2 (Leu) et 5 (Asp) ne peuvent pas être corrélées avec les monomères incorporés au sein de l'astine C (respectivement Pro et β -Phe). Néanmoins, il convient de rappeler que les prédictions réalisées sur les NRPSs d'origine fongique sont généralement peu fiables. En effet, comme décrit précédemment, les outils dédiés à l'analyse des NRPSs s'appuient sur des comparaisons avec des séquences déjà répertoriées dans les bases de données. Ces dernières sont majoritairement composées de séquences d'origine procaryote aux dépens des structures eucaryotes. Les faibles scores associés aux prédictions de NRPS Predictor2 semblent valider cette hypothèse.

3.3.2.2. Développement d'une méthode de prédiction alternative

Afin de développer un système de prédiction alternatif plus adapté à l'analyse des séquences fongiques, des travaux de modélisation de la synthétase AstS ont été menés. L'objectif consiste à obtenir une représentation tridimensionnelle de la NRPS afin de prédire la spécificité de substrat des domaines d'adénylation en se basant directement sur la conformation de la poche catalytique.

3.3.2.2.1. Recherche de templates appropriés

L'établissement d'un modèle s'appuie sur des alignements avec des matrices de référence, appelées « templates », dont la conformation 3D a déjà été expérimentalement élucidée. Une recherche des templates proches de la synthétase AstS a été réalisée via le logiciel HHPred. La taille des complexes NRPS et PKS étant un obstacle à la cristallisation, seuls des fragments de NRPS / PKS cristallisés sont disponibles en tant que template (**Tableau 5**).

Tableau 5 : Domaines / modules de PKS / NRPS cristallisés constituant des templates possibles pour la modélisation des différents modules de la NRPS AstS. (Structures d'origine eucaryote surlignées en bleu, et procaryote en blanc).

Type de domaines	Nom (PDB)	Organisme d'origine	NRPS/PKS associée	Spécificité de substrat
Module complet	5T3D	<i>Escherichia coli</i>	EntF	diOH-Bz
	4ZXH	<i>Acinetobacter baumannii</i>	AB3403	Gly
	2VSQ	<i>Bacillus subtilis</i>	SrfA-C	Leu
	5U89	<i>Geobacillus thermoglucosidasius</i>	DhbF	Thr
Domaine A + Domaine T	5ES6	<i>Brevibacillus parabrevis</i>	LgrA	Val
	5JNF	<i>Brevibacillus parabrevis</i>	LgrA	Val
	4DG8	<i>Pseudomonas aeruginosa</i>	PA1221	Val
	4IZ6	<i>Escherichia coli</i>	EntE/EntB	DHB
	4R0M	<i>Microcystis aeruginosa</i>	McyG	Phe-adenylate
Domaine T + Domaine C	2FQ1	<i>Escherichia coli</i>	EntB/EntE	DHB
	2JGP	<i>Brevibacillus parabrevis</i>	TycC	Leu
Zone inter-module	5U89	<i>Geobacillus thermoglucosidasius</i>	DhbF	-
	2JGP	<i>Brevibacillus parabrevis</i>	TycC	-
Domaine A	3VNR	<i>Streptomyces sp.</i>	CytC1	Abu
	5WMM	<i>Micromonospora sp.</i>	TioS4	N-Me-L-Cys
	4GR5	<i>Streptomyceslydicus</i>	SigN1deltaAsub	cPP
	1AMU	<i>Aneurinibacillus migulanus</i>	Gramicidine synthetase	Phe
	4OXI	<i>Vibrio cholerae</i>	AlmE	glycyl-adenosine-5-phosphate
	3O83	<i>Acinetobacter baumannii</i>	BasE	IXN
	5JJQ	<i>Streptomyces sp.</i>	IdnL1	3-aminobutanoic acid
	3E7W	<i>Bacillus subtilis</i>	DltA	D-Ala
	3WVN	<i>Streptomyceshalstedii</i>	VinN	L-aspartate
	1MDB	<i>Bacillus subtilis</i>	DhbE	DHB-adenylate
	5N81	<i>Brevibacillus parabrevis</i>	TycA	O-propargyl-beta-Tyr
	5KEI	<i>Mycobacterium smegmatis</i>	MbtA	acyl-adenylate
	3ITE	<i>Epichloe festucae</i>	SidN	cis-AMHO
Domaine T	2K0X	<i>Streptomyces coelicolor</i>	Actinorhodine synthase	-
	5TTB	<i>Yersinia pestis</i>	ArCP	-
	4NEO	<i>Streptomyces verticillus</i>	BlmI	-
	2MR7	<i>Actinoplanes teichomyceticus</i>	Teicoplanine synthase	-
	2JU1	<i>Saccharopolyspora erythraea</i>	6-deoxyerythronolideB	-
	2KR5	<i>Aspergillus parasiticus</i>	NSAS	-
Domaine C	1L5A	<i>Vibrio cholera</i>	VibH	-
	5DU9	<i>Streptomyces coelicolor</i>	CDA-C1	-
	5DLK	<i>Penicillium aethiopicum</i>	TqaA	-

Les organismes d'où proviennent ces NRPSs cristallisées sont très majoritairement procaryotes. La proportion de templates d'origine eucaryote est inférieure à 10 %.

Parmi les templates possibles identifiés, quatre présentent un score d'identité avec les modules de l'AstS supérieur ou égal à 20 % : 5T3D, 4ZXH, 2VSQ et 1AMU. Les templates 5T3D, 4ZXH et 2VSQ couvrent l'ensemble d'un module, tandis que 1AMU correspond à un domaine d'adénylation uniquement. L'utilisation des templates recouvrant l'ensemble d'un module est préférable pour modéliser les modules de l'AstS et pas uniquement les domaines isolés. Par ailleurs, les structures de 1AMU et du domaine d'adénylation de 5T3D présentent une conformation très similaire (Root Mean Square Deviation (RMSD) = 1.462 Å, superposition avec PyMOL). La structure tridimensionnelle des domaines d'adénylation semble très conservée entre les deux templates. Seules les chaînes correspondant à l'extrémité N-term du domaine se liant au domaine suivant présentent une flexibilité (**Figure 9A**).

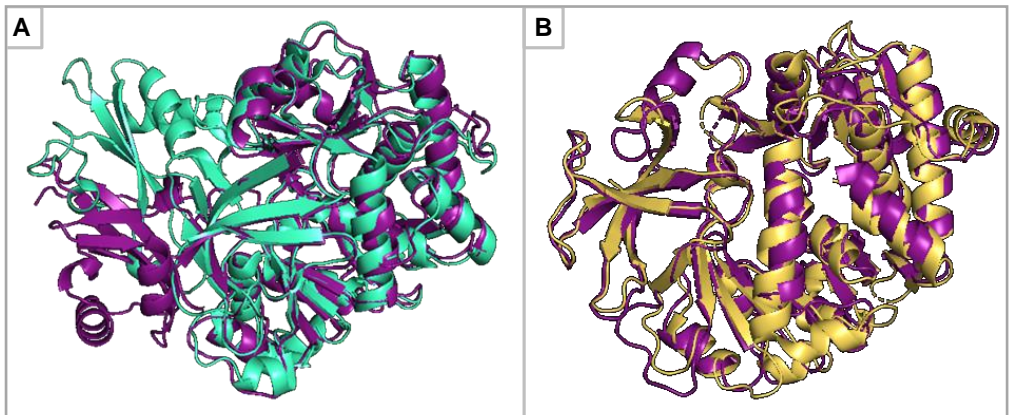


Figure 9 : Superposition des conformations tridimensionnelles de : **A**. 1AMU (violet) et 5T3D (cyan), RMSD = 1.462 Å ; **B**. 1AMU (violet) et 3ITE (jaune), RMSD = 1.275 Å.

La possibilité d'utiliser les templates d'origine fongique (3ITE, 2KR5, et 5DLK) a également été étudiée. Avec l'objectif de visualiser une éventuelle variation entre des domaines d'adénylation provenant de règnes différents, un alignement de 3ITE et 1AMU a été réalisé (**Figure 9B**). Le domaine d'adénylation fongique 3ITE présente une conformation similaire avec le domaine d'adénylation bactérien 1AMU (RMSD = 1.275 Å). Le positionnement des chaînes α et feuilletts β sont comparables entre les deux structures. L'alignement ne permet pas de mettre en évidence des différences structurales entre les 2 templates d'origines différentes. Les templates provenant d'organismes fongiques ne couvrant que des domaines isolés et présentant un score d'identité avec AstS moindre, 5T3D a été sélectionné en tant que template pour la modélisation de l'AstS. Le template 5T3D correspond au dernier module de l'entérobactine synthase EntF (C-A-T-Te) [30].

3.3.2.2. Génération d'un modèle tridimensionnel de la synthétase AstS

Les 5 modules composant AstS ont été modélisés individuellement via Phyre2. Une conformation similaire a été observée pour les 5 modules. Une représentation du modèle obtenu pour le module 2 est proposée **Figure 10**.

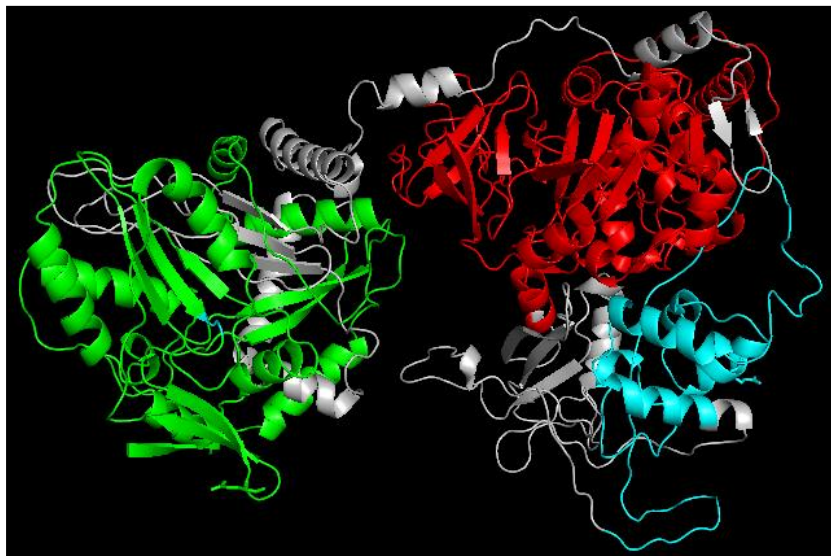


Figure 10 : Résultat de la modélisation du module 2 de l'AstS via le logiciel phyre2 (Visualisation PyMOL) avec en rouge le domaine d'adénylation, en bleu celui de thiolation, et en vert celui de condensation.

Les trois domaines (C-A-T) composant le module occupent chacun une zone spatiale propre. Les domaines d'adénylation (rouge) et de condensation (vert) présentent une structure tertiaire riche en hélices α et feuillets β . Le niveau de structuration de ces deux domaines est plus complexe que celui du domaine de thiolation (bleu) où très peu d'hélices α et feuillets β sont visibles.

3.3.2.2.3. Prédiction de la spécificité des domaines d'adénylation modélisés

Les modèles générés fournissent des informations concernant le dimensionnement de la poche catalytique des différents domaines d'adénylation de l'AstS. La conformation réelle de la poche catalytique doit permettre d'accueillir et de reconnaître spécifiquement le monomère à adényler (en présence d'ATP). Dans le cas de 1AMU, domaine d'adénylation cristallisé, les positions du monomère (Phe) et de l'AMP au moment de la réaction d'adénylation ont été caractérisées expérimentalement. Au sein de la poche catalytique de ce domaine A, le résidu Phe se positionne de façon que la fonction -OH du groupement carboxyle soit située à une distance de 3,7 Å de l'atome de phosphate du cofacteur AMP (**Figure 11A**).

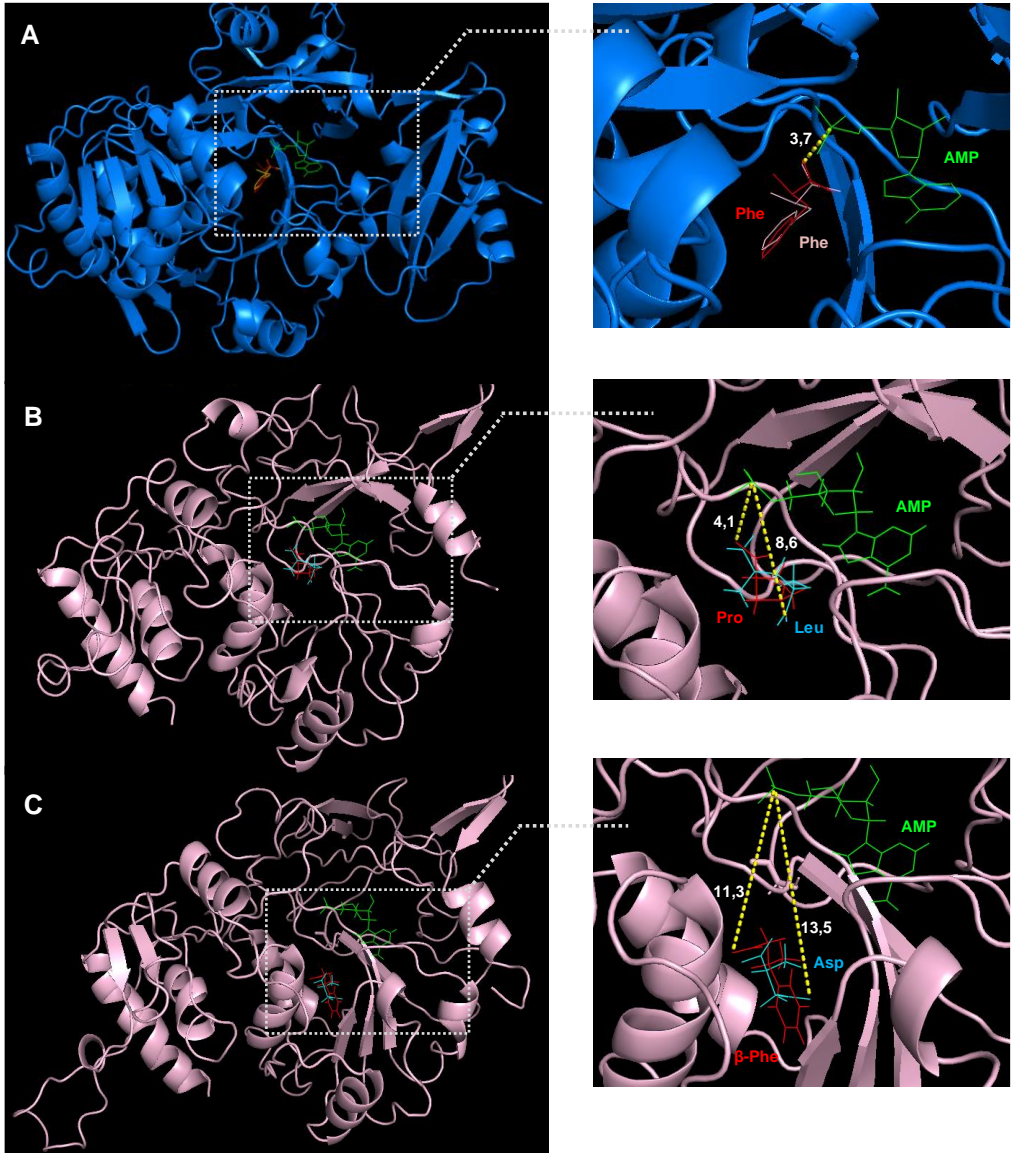


Figure 11 : Visualisation via PyMOL du positionnement des différents ligands au sein des domaines d'adénylation durant les simulations de « docking ». **A.** Validation de l'approche de docking sur la structure cristallisée de 1AMU en présence du cofacteur AMP. La position théorique du résidu Phe (en rose) (obtenu par « docking ») est comparée à la position expérimentale attribuée au substrat Phe (en rouge). **B.** Positionnement de l'AMP (vert), de la leucine (bleu) et de la proline (rouge) au sein du domaine d'adénylation du module 2 de l'AstS modélisée. **C.** Positionnement de l'AMP (vert), de l'Asp (bleu) et de la β -Phe (rouge) au sein du domaine d'adénylation du module 5 de l'AstS modélisée.

La faisabilité de l'approche de « docking » en tant que système de prédiction de la spécificité de substrat des domaines d'adénylation a, en premier lieu, été évaluée sur cette structure cristallisée. Un « docking » a donc été réalisé entre 1AMU (auquel le ligand Phe a été retiré) et les 20 acides aminés protéinogènes. Les scores obtenus pour chacun des résidus évalués sont présentés sur le **Tableau 6**.

Tableau 6: Scores obtenus pour le « docking » des 20 acides aminés protéinogènes au sein de 1AMU (domaine A cristallisé)

Acide aminé évalué	Score obtenu
Phénylalanine	54,69
Arginine	53,54
Tryptophane	49,75
Lysine	45,84
Histidine	45,57
Méthionine	44,19
Tyrosine	44,17
Leucine	40,93
Isoleucine	40,31
Glutamine	39,57
Acide glutamique	38,15
Thréonine	35,57
Valine	34,97
Asparagine	34,60
Sérine	33,50
Acide aspartique	32,95
Cystéine	32,52
Proline	31,61
Alanine	29,43
Glycine	28,50

Les scores obtenus sont inversement corrélés avec la quantité d'énergie nécessaire pour le passage de la forme libre du domaine d'adénylation à celle où le monomère se positionne au sein de la poche catalytique formant un complexe protéine-ligand [31,32]. Les valeurs indiquées pour les différents scores peuvent uniquement être utilisées pour réaliser des comparaisons entre les différents monomères testés dans des conditions identiques. Elles permettent de manière hypothétique, de classer les monomères selon leur degré d'affinité avec la poche catalytique du domaine A considéré. Dans le cas du domaine d'adénylation cristallisé 1AMU, cette approche a permis de prédire la phénylalanine comme étant le monomère le plus susceptible d'être incorporé (score le plus élevé). Ce monomère correspond à celui qui expérimentalement a été attribué au domaine d'adénylation 1AMU. Les positions de la phénylalanine prédite par « docking » et celle validée expérimentalement au sein de la poche catalytique sont similaires.

La même approche a été appliquée pour prédire la spécificité des domaines A de AstS à partir du modèle tridimensionnel généré. Les scores de « docking » au sein des poches catalytiques ont été calculés pour chacun des monomères attendus pour l'assemblage des astines (dans l'ordre des modules : L-2Abu, L-Pro, L-2Abu, L-Ser, β -Phe). Ces scores ont ensuite été comparés avec ceux obtenus pour le « docking » des monomères prédits par NRPS Predictor 2 (dans l'ordre des modules : Val, Leu, Gln, Cys, Asp). Les monomères prédits par NRPS Predictor 2 ont obtenu le score le plus favorable pour le « docking » au sein des modules 1, 2, et 3, tandis que ceux initialement attendus pour l'assemblage des astines ont généré un meilleur score au sein des modules 4 et 5.

Une évaluation de la position adoptée par les différents ligands testés au sein des domaines A (en présence d'AMP) a été réalisée. À titre d'exemple, le positionnement obtenu au sein du domaine d'adénylation des modules 2 et 5 est présenté respectivement sur la **Figure 11B** et la **Figure 11C**. La position du ligand Pro vis-à-vis de l'AMP (4,1 Å) semble en adéquation avec la distance observée chez 1AMU. En revanche, l'orientation du ligand Leu (qui a obtenu le score le plus élevé) entraîne un éloignement marqué du groupement carboxyle vis-à-vis de l'atome phosphore de l'AMP (8,6 Å). Ce positionnement du groupement carboxyle ne permet pas à la réaction d'adénylation de s'établir. Une observation similaire a été réalisée pour certains des ligands testés au sein des domaines A₁, A₃ et A₄. En ce qui concerne le module 5, les monomères évalués sont tous deux positionnés de manière distante vis-à-vis de l'AMP (11,3 Å pour la β -Phe et 13,5 Å pour l'Asp) (**Figure 11C**). Cet éloignement prononcé empêche la réaction d'adénylation de s'établir entre ces deux structures.

La position de la plupart des ligands testés au sein des domaines A modélisés de l'AstS n'est donc pas en adéquation avec le positionnement requis pour que la réaction d'adénylation s'établisse. Cet élément biaise les scores obtenus en termes de « docking » avec la structure modélisée. Le « docking » positionne les ligands et l'AMP de manière à obtenir la conformation de plus basse énergie. Ce positionnement tient compte de l'encombrement stérique et des différentes charges, mais n'intègre pas les éventuelles réactions intervenantes entre les ligands. Ces éléments doivent être pris en compte pour nuancer les prédictions obtenues via un « docking » des ligands au sein de domaines d'adénylation modélisés. Dans le cas d'une structure cristallisée, comme 1AMU, la position exacte des monomères et de l'AMP est déterminée expérimentalement alors que dans les simulations de docking, les groupements carboxyle et phosphate sont éloignés car ils présentent des charges identiques.

4. Discussion

C. asteris est un champignon endophyte qui a été isolé très récemment [4]. Les 38 clusters de gènes identifiés comme étant impliqués dans la biosynthèse de métabolites secondaires témoignent de son fort potentiel en tant que producteur de composés bioactifs d'intérêt. L'exploitation et l'étude de ce potentiel n'en est qu'à ses prémices. Seul l'un des clusters, relatif à une structure PKS, a été formellement attribué à la voie de biosynthèse de l'anthraquinone dimérique skyrin. Ce composé constitue le principal pigment responsable de la coloration atypique de *C. asteris* [6]. En ce qui concerne les autres clusters, aucune étude à ce jour, n'a pu établir de connexion avec un composé produit. Le cluster n° 15 (**Tableau 2**), s'il n'est pas cryptique, code pour une synthétase de 4 modules avec une structuration particulière (A-T-C-A-T-C-T-C- C_{term}). Les modules 3 et 4 comportent uniquement des domaines T et C et sont dépourvus de domaines d'adénylation. Cette structuration est typique d'un mode de fonctionnement itératif de la NRPS, où un même monomère activé par les domaines A_1 et A_2 est incorporé à plusieurs reprises dans le NRP sous l'action des modules 3 et 4. Ce mode de fonctionnement itératif a par exemple été décrit pour les NRPSs de la beauvericine et de la bassianolide chez *Beauveria bassiana* [33], ou pour la NRPS du ferrichrome chez *Schizosaccharomyces pombe* [34]. La synthétase codée par le cluster n° 6 présente une structure appropriée pour l'assemblage linéaire d'un pentapeptide cyclique tel que l'astine C. En effet, celle-ci comporte 5 modules comprenant chacun un domaine d'adénylation et s'achève par un domaine de condensation terminal.

L'analyse des domaines de condensation de la NRPS codée par le gène *astS* présent dans ce cluster n° 6, a permis de mettre en évidence les spécificités du domaine de condensation terminal par rapport aux quatre autres. En effet, celui-ci est plus proche en termes de séquences des domaines C_{term} des NRPSs Cct et HCct de *T. islandicus*, ainsi que de *cyclo1_C12_dual*, que des quatre autres domaines C de l'AstS. Le domaine de condensation *cyclo1_C12* est un domaine C_{term} de la cyclosporine synthétase, dont l'implication dans la cyclisation du NRP a été démontrée [35]. Par conséquent, en prenant en compte ces similitudes avec trois autres C_{term} impliqués dans la cyclisation de NRP, on peut déduire que le C_{term} de l'AstS présente une activité de cyclisation similaire. En revanche, les quatre premiers domaines de condensation de l'AstS ont été identifiés comme des sous-domaine « dual C/E » par le logiciel NaPDoS. Ce sous-type de domaine C est caractéristique des domaines de condensation composant une synthétase qui incorpore des acides aminés sous leur forme D au sein du NRP [28]. À l'inverse, cette classification n'est pas attendue pour un NRP composé uniquement de monomère sous la forme L, comme c'est le cas pour l'astine C. De même, l'analyse des spécificités de substrat des domaines d'adénylation, via NRPS predictor2, a fourni des résultats contrastés. Aucun monomère incorporé dans l'astine n'a été prédit au niveau « Nearest neighbor ». Les monomères L-2Abu ont en revanche été proposés comme substrat possible pour les domaines A_1 et A_3 à l'échelle « Large cluster ». Par ailleurs, le substrat L-Cys, dont la structure est très proche de la L-Ser, a été proposé comme substrat pour le domaine A_4 . Les autres domaines analysés n'ont pas pu être corrélés avec l'un des monomères

incorporés dans l'astine C. Dans les faits, les analyses bioinformatiques réalisées sur des NRPSs d'origine fongique sont souvent peu fiables. Il existe des différences structurales entre les domaines provenant de NRPSs fongiques et celles d'origine bactérienne. Les outils bioinformatiques actuels, basés essentiellement sur des alignements et des comparaisons avec des séquences déjà répertoriées dans les bases de données, ne tiennent pas compte de ces spécificités. Les séquences relatives aux NRPSs d'origine bactérienne étant majoritaires au sein des bases de données utilisées, l'analyse des NRPSs fongiques peut être biaisée [11].

La modélisation de la structure tridimensionnelle des modules de la synthétase supposée d'astine permet de visualiser la conformation adoptée par les différents domaines. La structure tertiaire des domaines de thiolation est plus relâchée, ce qui laisse supposer une certaine flexibilité, indispensable pour présenter les monomères activés par les domaines A aux domaines C [36]. Le modèle de la synthétase AstS qui a été généré a ensuite été utilisé afin de développer une approche alternative de prédiction de la spécificité de substrat des domaines A. Cette approche consiste à déterminer, en fonction de la dimension de la poche catalytique, le monomère le plus susceptible d'être reconnu et activé. Le positionnement des monomères vis-à-vis de l'AMP, requis lors de l'adénylation, dans la poche catalytique des domaines A modélisés s'est avéré problématique. L'orientation du groupement carboxyle de l'acide aminé par rapport au groupement phosphate de l'AMP n'est pas celle observée expérimentalement au sein des domaines A au moment de la réaction d'adénylation [37]. Deux éléments peuvent expliquer cette orientation inappropriée : (1) le positionnement approximatif de l'AMP estimé à partir des observations de 1AMU, et (2) les charges identiques portées par le groupement carboxyle du monomère et le phosphate de l'AMP. Les scores obtenus à partir de ce « docking » sont donc à considérer avec circonspection, d'autant plus que les ligands testés ont des propriétés stéréochimiques proches, ce qui complique leur classement par le programme de docking. Couplé au fait que le docking est réalisé sur des modèles par homologie et non des structures résolues, cette approche semble inadaptée dans l'état actuel pour la prédiction de ligands. Afin d'obtenir des données de prédictions davantage exploitables, il conviendrait de considérer des structures déjà résolues et cristallisées, comme l'a montré le docking d'acides aminés dans 1AMU.

Malgré les problématiques relatives à l'analyse de génomes fongiques comme celui de *C. asteris*, plusieurs éléments permettent d'associer la synthétase AstS et le cluster n° 6 à la biosynthèse de l'astine. Tout d'abord, AstS est la seule synthétase possédant 5 modules, capable d'assembler un pentapeptide de manière linéaire, à avoir été détectée lors de l'analyse du génome de *C. asteris*. Les prédictions de spécificité de substrat des domaines d'adénylation, bien que peu fiables, semblent indiquer que L-2Abu est un monomère possible pour A₁ et A₃, et que L-Ser pourrait être reconnu par A₄. La NRPS s'achève par un domaine de condensation terminal, présentant une forte homologie de séquence avec les domaines C_{term} des NRPSs de la Cct, l'HCct et la cyclosporine. Ce domaine C_{term} est probablement impliqué dans la cyclisation du NRP assemblé, comme cela a été démontré pour les C_{term} homologues [7,35]. Par ailleurs, le cluster n° 6 comporte plusieurs gènes codant pour des protéines potentiellement

impliquées dans la voie de biosynthèse de l'astine. En particulier, certains gènes comme *astU* et *astV* codent pour des enzymes impliquées respectivement dans la biosynthèse de β -Phe et L-2Abu, des monomères précurseurs de l'astine C. Le gène *astT* est associé à des transporteurs membranaires de type ABC, potentiellement impliqués dans le transport et la sécrétion du peptide d'astine C. En combinant ces différents éléments, et en considérant la très forte similitude de structure entre la NRPS AstS et celle de Cct identifiée chez *T. islandicus*, une représentation schématique de la voie de biosynthèse de l'astine a pu être établie (**Figure 12**).

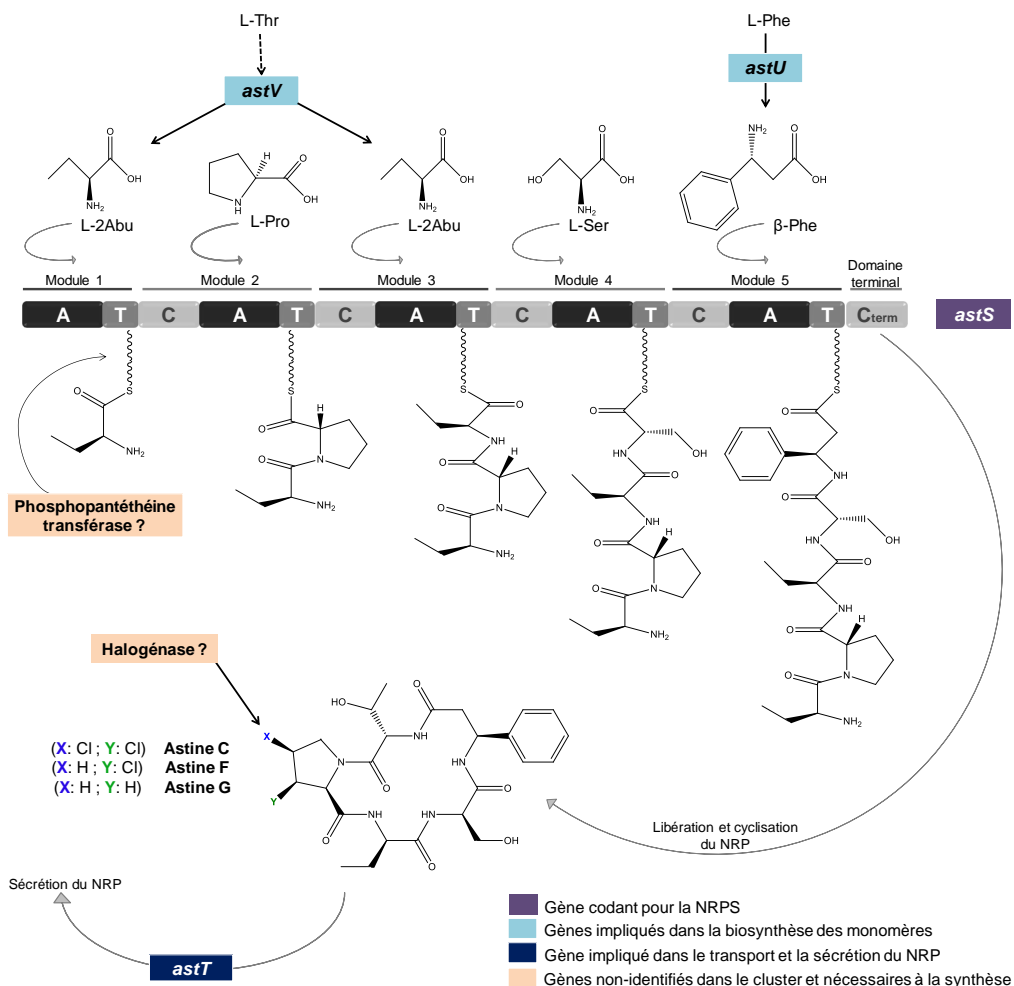


Figure 12 : Proposition de schéma détaillant la voie de biosynthèse des astines C, F et G. La synthèse de 5 modules est codée par le gène *astS*, puis activée par l'action d'une phosphopantéthéine transférase dont le gène n'a pas été identifié. Les gènes *astU* et *astV* sont potentiellement impliqués dans la synthèse des monomères précurseurs de l'astine (respectivement β -Phe et L-2Abu). Une fois assemblé, le peptide est libéré et cyclisé par l'action du domaine C_{term}, générant ainsi le NRP d'astine sous la forme G. Les formes d'astine F et C (mono- et di-chlorées) sont éventuellement obtenues par l'action d'une halogénase qui n'a pas été identifiée au sein du cluster de gènes n° 6. Le gène *astT* (transporteur de type ABC) est ensuite potentiellement impliqué dans la sécrétion du NRP.

L'ordre d'assemblage des monomères composant l'astine par les différents modules est suggéré comme suit (du module 1 au module 5) : L-2Abu, L-Pro, L-2Abu, L-Ser, β -Phe. Cette disposition est en adéquation avec les éléments de prédiction obtenus lors de l'analyse de la NRPS AstS, et présente une forte similitude avec l'ordre observé chez la NRPS de la Cct (L-Ser, L-Pro, L-2Abu, L-Ser, β -Phe), homologue à 62 % avec AstS [7]. Certains éléments, pourtant indispensables à la formation d'astine C, n'ont pas été mis en évidence au sein du cluster n° 6. En particulier, aucun gène codant pour une phosphopantéthéine transférase permettant l'activation de la synthétase AstS n'a été détecté. De même, une halogénase, dont le gène n'a pas été identifié, est potentiellement impliquée dans la conversion de l'astine G (non-chlorée) en astine F et C (respectivement mono- et di-chlorée). Une perspective à explorer en termes d'analyses bioinformatiques consistera à identifier, au sein du génome de *C. asteris*, ces deux gènes manquants.

Le gène *astS* (17 kb), évalué au cours de cette étude, constitue l'élément central de la voie de biosynthèse de l'astine, puisqu'il code pour la synthétase capable d'assembler le pentapeptide cyclique (**Figure 3**). Étant donné son importance dans la biosynthèse de l'astine, et sa longueur très peu usuelle, il convient de considérer en priorité le transfert de ce gène lors de la mise en place d'un système de production de l'astine chez un hôte hétérologue.

5. Références

1. Kosemura, S., et al. (1993) Isolation and Structure of Asterin, a new halogenated cyclic pentapeptide from *Aster tataricus*. *Tetrahedron Lett.*, **34** (8), 1291–1294.
2. Cheng, D.L., et al. (1996) New pentapeptides from *Aster tataricus*. *Phytochemistry*, **41** (1), 225–227.
3. Morita, H., et al. (1993) Astins A and B, antitumor cyclic pentapeptides from *Aster tataricus*. *Chem. Pharm. Bull. (Tokyo)*, **41** (5), 992–993.
4. Jahn, L., et al. (2017) *Cyanodermella asteris* sp. nov. (Ostropales) from the inflorescence axis of *Aster tataricus*. *Mycotaxon Issn*, **132** (March), 107–123.
5. Jahn, L. (2015) Characterization of a new endophytic astin producer, *Pelliciarosea asterica*, from *Aster tataricus*. *PhD Thesis*, (October 2015), 1–154.
6. Jahn, L., et al. (2017) Linking secondary metabolites to biosynthesis genes in the fungal endophyte *Cyanodermella asteris*: The anti-cancer bisanthraquinone skyrin. *J. Biotechnol.*, **257** (December 2016), 233–239.
7. Schafhauser, T., et al. (2016) The cyclochlorotine mycotoxin is produced by the nonribosomal peptide synthetase CctN in *Talaromyces islandicus* ('*Penicillium islandicum*'). *Environ. Microbiol.*, **18** (11), 3728–3741.
8. Leclère, V., et al. (2016) Bioinformatics Tools for the Discovery of New Nonribosomal Peptides, in *Methods in molecular biology (Clifton, N.J.)*, vol. 1401, pp. 209–232.
9. Weber, T., and Kim, H.U. (2016) The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.*, **1** (2), 69–79.
10. Minowa, Y., et al. (2007) Comprehensive Analysis of Distinctive Polyketide and Nonribosomal Peptide Structural Motifs Encoded in Microbial Genomes. *J. Mol. Biol.*, **368** (5), 1500–1517.
11. Nielsen, J.C., and Nielsen, J. (2017) Development of fungal cell factories for the production of secondary metabolites: Linking genomics and metabolism. *Synth. Syst. Biotechnol.*, **2** (1), 5–12.
12. Blin, K., et al. (2017) AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.*, **45** (W1), W36–W41.
13. Röttig, M., et al. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.*, **39** (suppl_2), W362–W367.
14. Söding, J., et al. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33** (Web Server issue), W244–248.
15. Kelley, L.A., et al. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10** (6), 845–58.
16. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.
17. Laskowski, R.A., et al. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26** (2), 283–291.
18. Liebeschuetz, J.W., et al. (2012) Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comput. Aided. Mol. Des.*, **26** (6), 737–48.
19. Gao, X., et al. (2012) Cyclization of Fungal Nonribosomal Peptides by a Terminal Condensation-Like Domain. *Nat Chem Biol*, **412** (2), 426–434.
20. Sekowska, A., et al. (2004) Bacterial variations on the methionine salvage pathway. *BMC Microbiol.*, **4**, 9.
21. Musacchio, A., et al. (1992) SH3 - an abundant protein domain in search of a function. *FEBS Lett.*, **307** (1), 55–61.
22. Rodaway, A.R., et al. (1989) Similarity in membrane proteins. *Nature*, **342** (6250), 624.

23. Ravid, T., and Hochstrasser, M. (2008) Diversity of degradation signals in the ubiquitin-proteasome system. *Nat. Rev. Mol. Cell Biol.*, **9** (9), 679-690.
24. Amerik, A.Y., and Hochstrasser, M. (2004) Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta - Mol. Cell Res.*, **1965** (1-3), 189-207.
25. Ding, H., et al. (2017) Inhibitors of BRD4 protein from a marine-derived fungus *Alternaria* sp. NH-F6. *Mar. Drugs.*, **15** (3).
26. Mietton, F., et al. (2017) Selective BET bromodomain inhibition as an antifungal therapeutic strategy. *Nat. Commun.*, **18** (8).
27. Mohamed A. Marahiel, et al. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.*, **97** (7), 2651-2674.
28. Walton, J.D. (2006) HC-toxin. *Phytochemistry.*, **67** (14), 1406-1413.
29. Caradec, T., et al. (2014) Prediction of monomer isomery in florine: A workflow dedicated to nonribosomal peptide discovery. *PLoS One*, **9** (1).
30. Drake, E.J., et al. (2016) Structures of two distinct conformations of *holo*-non-ribosomal peptide synthetases. *Nature*, **529** (7585), 235–258.
31. Ramírez, D., and Caballero, J. (2018) Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules*, **23** (5).
32. Kollman, P.A., et al. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **33** (12), 889–897.
33. Yu, D., et al. (2017) Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.*, **8** (May), 1–11.
34. Schwecke, T., et al. (2006) Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. *ChemBioChem*, **7** (4), 612–622.
35. Zhang, J., et al. (2016) Structural basis of nonribosomal peptide macrocyclization in fungi. *Nat. Chem. Biol.*, **12** (12), 1001–1003.
36. Winn, M., et al. (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.*, **33** (2), 317–347.
37. Challis, G.L., and Naismith, J.H. (2004) Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Struct. Biol.*, **14** (6), 748–756.

Chapitre 4

**Mise en place d'un système d'expression
hétérologue chez la levure**

1. Introduction

Le nombre de génomes séquencés disponibles au sein des bases de données n'a cessé d'augmenter ces dernières années [1]. Les outils bioinformatiques développés permettent de les analyser afin de mettre en évidence la production potentielle de composés d'intérêts [2]. Les taux de production observés chez les organismes naturellement producteurs sont bien souvent très faibles voire nuls. Les conditions de culture de ces micro-organismes néo-séquencés sont souvent mal maîtrisées, et l'absence d'outils moléculaires dédiés limite les possibilités de manipulations génétiques. Le transfert de la voie de biosynthèse du composé d'intérêt au sein d'un hôte hétérologue est une approche courante pour remédier à ce problème [3].

Cyanoderrella asteris, le producteur naturel de l'astine C, présente un taux de croissance très faible sur les milieux de culture conventionnellement utilisés pour les champignons filamenteux (PDA, MEA). Ce faible taux de croissance limite fortement la production du peptide non-ribosomique d'astine C [4]. L'utilisation combinée de diverses méthodes et outils bioinformatiques a permis d'identifier le cluster de gènes potentiellement impliqué dans la biosynthèse de l'astine (**Figure 3 – Chapitre 3**). Ces données peuvent être utilisées pour envisager le transfert de la voie de biosynthèse dans un hôte hétérologue présentant des conditions de culture plus adaptées à la production du NRP.

Le choix de l'hôte est un paramètre très important en matière de production hétérologue. La proximité taxonomique entre l'organisme d'origine et l'hôte choisi est un facteur à prendre en compte pour limiter les problématiques liées à l'usage des codons. Les levures, taxonomiquement proches des champignons filamenteux, sont particulièrement adaptées pour réaliser une expression hétérologue des gènes fongiques. De plus, pour bon nombre de levures, une grande diversité d'outils moléculaires sont disponibles, et leurs conditions de cultures sont bien maîtrisées [5]. La levure *Saccharomyces cerevisiae* a été sélectionnée à de nombreuses reprises pour produire de manière hétérologue des NRPs (**Tableau 2 – Chapitre 1**). La capacité de cette levure à assembler efficacement de longues séquences par TAR Cloning facilite le clonage des gènes codants pour les NRPSs [6]. La levure *Hansenula polymorpha* a également été utilisée pour la production hétérologue de l'ACV, un NRP précurseur de la pénicilline [7]. La souche *Schizosaccharomyces pombe*, en tant que seule levure capable de produire de manière homologue un NRP (le ferrichrome), est également une souche envisageable pour l'expression hétérologue de NRPs [8]. Cependant, *S. pombe* n'étant pas initialement considérée comme une souche à haut potentiel de production, peu d'outils moléculaires ont été développés afin de faciliter sa transformation [9]. À l'inverse, les levures *Pichia pastoris* et *Yarrowia lipolytica* disposent de nombreux outils moléculaires et d'une remarquable capacité de production de métabolites secondaires. La voie de biosynthèse du polykétide citrinine a été reconstituée chez *P. pastoris* [10], tandis que *Y. lipolytica* a été exploitée notamment pour la production de divers terpénoïdes [11,12]. Bien que ces levures n'aient jamais été utilisées pour exprimer des NRPSs, leur aptitude à se développer à très haute densité cellulaire, et l'existence de promoteurs inductibles permettant

d'obtenir de hauts taux de transcription, en font des hôtes potentiels pour mettre en place une production hétérologue de NRPs [13].

Dans ce chapitre, deux approches ont été développées parallèlement. Dans un premier temps, il s'agit de mettre à profit la capacité de *S. cerevisiae* à assembler de grandes séquences par TAR cloning, pour exprimer l'intégralité du gène codant pour l'astine synthétase (**Figure 12 – Chapitre 3**). La seconde consiste à développer pour la première fois une production hétérologue de NRPs chez *Y. lipolytica*, une levure à haut potentiel de production, encore inexploitée en termes de NRPs. Une NRPS de taille plus réduite sera clonée chez cette levure, afin d'évaluer sa capacité à exprimer une séquence NRPS et à exploiter la voie de production non-ribosomique. Par conséquent, une construction génétique chimérique sera générée, encodant une synthétase tronquée se composant des deux premiers modules associés au domaine de condensation terminal (**Figure 1**).

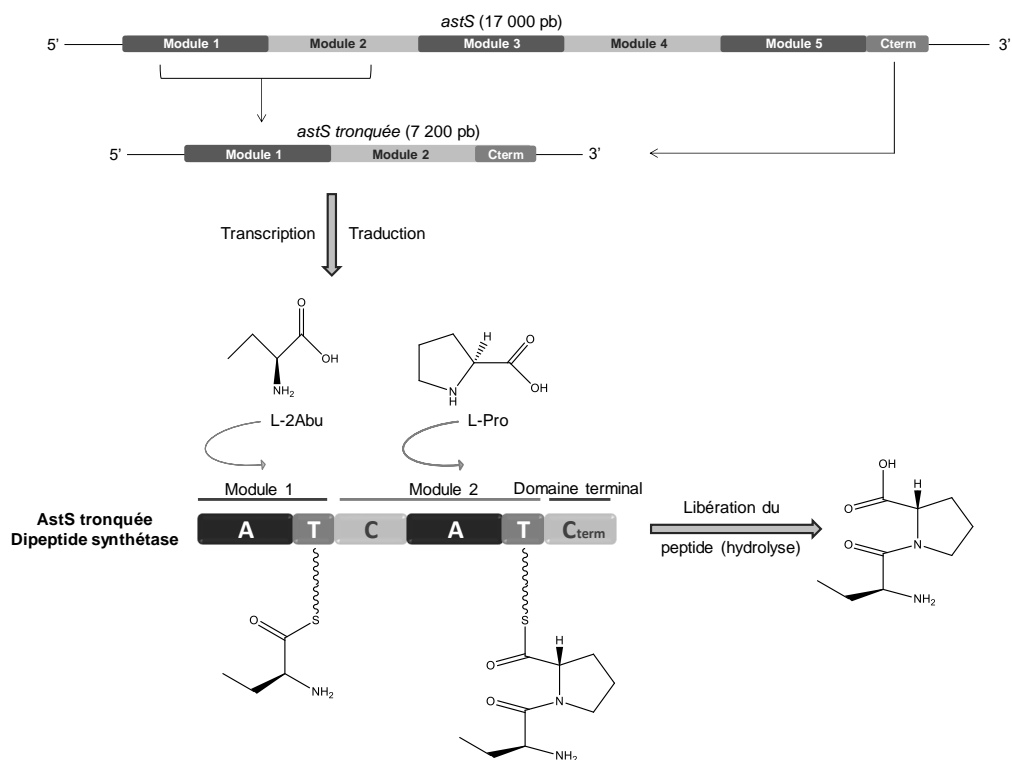


Figure 1 : Séquence chimérique générée pour le clonage dans *Y. lipolytica*, composée d'un fragment de la séquence de *astS* codant pour les deux premiers modules, directement suivis, dans le même cadre de lecture ouvert (ORF), par la séquence codant pour le domaine de condensation terminal. La synthétase tronquée produite, si elle est fonctionnelle, doit assembler et libérer un dipeptide L-2Abu – L-Pro.

2. Matériels et méthodes

2.1. Souches utilisées

2.1.1. Souches de champignons filamenteux

Le champignon filamenteux *Cyanoderrella asteris* (Mycobank #814158) a été isolé à partir des racines d'*Aster tataricus* d'après le protocole décrit par Jahn *et al.* (2015) [4].

La souche *Aspergillus nidulans* (IHEM 22446), provenant de "BCCM/IHEM Biomedical Fungi and Yeasts Collection", a été utilisée au cours de cette étude afin d'isoler le gène *npgA* codant pour une phosphopantéthéine transférase.

2.1.2. Souches de levures

Les travaux relatifs à *Saccharomyces cerevisiae* ont été réalisés avec la souche DSy5 (*MAT α leu2 trp1 ura3-52 his3 ::GALI-GAL4 pep4 prb1-1122*), utilisée précédemment à de multiples reprises pour la production de protéines recombinantes [14–16]. Cette souche est auxotrophe pour la leucine, le tryptophane, l'uracile, et l'histidine. Elle est également mutée pour les séquences activatrices de protéases (*pep4 prb1-1122*).

La souche RIY261 (*MatA, leu2-, ura3-, xpr2-, Δ EYD1*), dérivée de la souche *Y. lipolytica* Po1d (*MatA, leu2-, ura3-, xpr2-*) [17] présente une auxotrophie pour la leucine et l'uracile. Cette souche est mutée pour le gène *EYD1* codant pour une erythritol déshydrogénase qui catalyse l'oxydation de l'erythritol en erythrulose [18]. La souche RIY261 n'est donc pas capable de convertir l'erythritol synthétisé ou prélevé dans le milieu.

2.1.3. Souches de bactéries

Le maintien des constructions plasmidiques a été réalisé au sein de la souche *Escherichia coli* DH5 α . La souche *E. coli* One Shot[®] OmniMAX[™] (Invitrogen, Carlsband, États-Unis) a été utilisée pour produire les plasmides pESC-ura-sfp-AstS.PCR, pESC-ura-sfp-AstS.RESTR, et pYX112-AstS.RESTR, afin d'obtenir une quantité suffisante pour permettre leur séquençage.

La souche de *Bacillus subtilis* BBG111 décrite par Coutte *et al.* (2010) [19], dérivée de la souche *B. subtilis* 168, a été utilisée afin d'isoler le gène *sfp* codant pour une phosphopantéthéine transférase fonctionnelle.

2.2. Conditions de cultures

2.2.1. Culture des champignons filamenteux

C. asteris et *A. nidulans* ont été cultivés à 30 °C sur milieu gélosé MEA (extrait de malt 30 g/L, glucose 20 g/L, peptone extrait de viande 1 g/L, agar 20 g/L). Les cultures de *C. asteris* et d'*A. nidulans* ont été réalisées respectivement pendant 42 et 10 jours. Les différents composants utilisés dans les milieux de culture ont été fournis par Sigma-Aldrich (St. Louis, États-Unis).

2.2.2. Culture des bactéries

Les souches *E. coli* et *B. subtilis* utilisées ont été cultivées à 37 °C, avec agitation à 160 rpm, sur milieu LB (« Lysogeny broth », tryptone 10 g/L, extrait de levure 5 g/L, chlorure de sodium 10 g/L). Dans le cas de milieu gélosé, 15 g d'agar ont été ajoutés par litre de milieu.

Les milieux utilisés pour sélectionner les transformants positifs d'*E. coli* ont été supplémentés avec un antibiotique (ampicilline ou kanamycine) à une concentration finale de 100 µg/mL.

2.2.3. Culture des levures

Le milieu YPD (« Yeast extract, peptone, dextrose », extrait de levure 10 g/L, peptone bactériologique 20 g/L, glucose 20 g/L) a été utilisé pour conserver les souches de *S. cerevisiae* et *Y. lipolytica* non-transformées. Dans le cas de milieu gélosé, 15 g d'agar ont été ajoutés par litre de milieu.

Les souches *S. cerevisiae* et *Y. lipolytica* transformées ont été sélectionnées puis maintenues sur un milieu YNB - Yeast Nitrogen Base Without Amino Acids – (Sigma-Aldrich, St Louis, États-Unis). Selon l'auxotrophie de la souche considérée, un ou plusieurs acides aminés ont été ajoutés au milieu dans les concentrations suivantes : uracile (20 µg/mL), leucine (40 µg/mL), lysine (40 µg/mL), histidine (20 µg/mL), tryptophane (30 µg/mL).

Dans le cadre des analyses relatives à la transcription et la traduction de la séquence chez *S. cerevisiae*, les transformants DS.Pe.Ec73 / DS.Py.62, et la souche non-transformée DSy5, ont été cultivés à 30 °C dans 200 mL de milieu SD Medium-URA (MP-Biomedical, Solon, États-Unis), sous agitation à 160 rpm. Le milieu utilisé pour les cultures de DS.Pe.Ec73 (et DSy5, contrôle négatif) a été supplémenté avec 10 g/L de galactose pour permettre une induction des promoteurs GAL1 / GAL10. Les souches ont été inoculées à une DO₆₀₀ de 0,1.

Dans le cadre des analyses relatives à la transcription et la traduction des séquences chez *Y. lipolytica*, les transformants B3 / B3R2, et la souche non-transformée RIY261, ont été cultivés à 30 °C dans 200 mL de milieu, sous agitation à 160 rpm. Deux milieux différents ont été utilisés : le milieu YPD ou le milieu YNB (supplémenté en Leu / Ura pour RIY261, et en Leu pour B3). Pour l'ensemble des cultures réalisées, les milieux ont été supplémentés avec de l'erythritol (10 g/L) afin d'induire l'expression du promoteur pEYK300A3B. Les souches ont été inoculées à une DO₆₀₀ de 0,1.

2.3. Extraction des ADN génomiques

2.3.1. Méthode classiquement utilisée

L'extraction de l'ADN génomique des organismes fongiques (levures et champignon filamenteux) a été réalisée en adaptant partiellement la méthode décrite par Raeder et Broda (1995) [20]. Les cellules fongiques ont été mises en suspension dans un tampon d'extraction (200 mM Tris HCl pH 8,5, 250 mM NaCl, 25 mM

EDTA, 0,5 % SDS), auquel un volume équivalent de phénol-chloroforme (70 %-30 %) a été ajouté. Après centrifugation (30 min – 11000 × g), la phase aqueuse a été récupérée, puis l'ARN dégradé par traitement à la RNase A. La solution a été traitée avec un volume équivalent en chloroforme puis centrifugée (15 min – 11000 × g). La fraction supérieure contenant l'ADN a été conservée, et un ½ volume d'isopropanol y a été ajouté, entraînant la précipitation de l'ADN. Le culot d'ADN a été lavé par une solution d'éthanol (99 %) puis séché à température ambiante pour finalement être remis en suspension dans une solution aqueuse. La quantité d'ADN extrait a été estimée par dosage spectrophotométrique (Nanodrop 2000®, Wilmington, États-Unis).

2.3.2. Mise au point d'une méthode d'extraction pour l'ADN de *C. asteris*

L'extraction de l'ADN de *C. asteris* a été réalisée à partir de cellules fongiques (100 mg ± 20 mg) prélevées à partir de cultures âgées de 42 jours sur milieu MEA gélosé. Trois conditions de pré-traitement lytique ont été considérées : (1) broyage à l'azote liquide, (2) lyse en présence de billes de verre (0,10 mm de diamètre) au dismembrateur (Mikro-Dismembrator, Sartorius, Allemagne), (3) une combinaison de ces deux traitements (1 puis 2). Dans le cas de la lyse via dismembrateur, les cellules ont été préalablement mises en suspension dans un tampon d'extraction provenant du kit utilisé pour la purification de l'ADN à l'issue de la lyse. La lyse a été effectuée via 5 cycles de 2 min à 2000 rpm, alternés de 2 min où l'échantillon était maintenu sur glace. Les billes de verre et les cellules ont été éliminées par centrifugation (10 min, 10 000 × g, 4 °C), et le surnageant conservé et traité selon les conditions préconisées par les différents kits testés.

Plusieurs kits ont ensuite été évalués pour l'extraction et la purification de l'ADN génomique à partir du lysat cellulaire de *C. asteris* : Qiagen DNeasy kit (Qiagen Inc., Valencia, États-Unis), Norgen soil DNA isolation kit, et Norgen Plant / Fungi DNA isolation kit (Norgen, Thorold, Canada). La quantité d'ADN extrait, estimée par dosage spectrophotométrique, est présentée en ng/mg de cellules fongiques considérées pour l'extraction.

2.4. Isolement du gène codant pour l'astine synthétase

2.4.1. Par restriction de l'ADN génomique

Une restriction de l'ADN génomique de *C. asteris* a été réalisée afin d'obtenir des fragments d'ADN (dont celui contenant *astS*) pour la transformation chez *S. cerevisiae*. Deux protocoles de restriction enzymatique ont été développés. Dans le premier cas 43 µL d'ADNg (~20 ng/µL) ont été ajoutés à un mélange contenant 5 µL de tampon « R » et 2 µL de *SgrDI* (10 unités) (Thermo Fisher Scientific, Boston, États-Unis). Le mélange a été incubé à 37 °C pendant 24 h, puis purifié via le kit GeneJET PCR Purification Kit (Thermo Fisher Scientific, Boston, États-Unis) avant d'être transformé chez *S. cerevisiae*. Pour le second protocole, 35 µL d'ADNg (~20 ng/µL) ont été ajoutés à un mélange contenant 5 µL de tampon SuRE / Cut H, 9 µL H₂O, et 1 µL de *EcoRI* (10 unités) (Roche Applied Science, Basel, Suisse). Le

mélange a été incubé 5 min à 37 °C, puis 15 min à 65 °C afin d'inactiver l'enzyme. Les fragments d'ADN restreints ont ensuite été purifiés via le kit GeneJET PCR Purification Kit avant d'être transformés chez *S. cerevisiae*.

2.4.2. Par amplification PCR

Plusieurs ADN polymérases ont été testées pour l'amplification du gène *astS* de 17 kb en utilisant le couple d'amorces *AstS1* / *AstS2*. La Taq Phusion™ HF DNA (Thermo Fisher Scientific, Boston, États-Unis) a été évaluée dans un mélange contenant 1 µL d'ADNg (~20 ng/µL), 10 µL de tampon Phusion HF, 1 µL de dNTPs (10 mM), 2,5 µL d'amorce fw (« forward » - sens, 10 µM), 2,5 µL d'amorce rv (« reverse » - antisens, 10 µM), 32,5 µL H₂O, 0,5 µL de Phusion DNA (1 unité). La DreamTaq Green DNA polymérase (Thermo Fisher Scientific, Boston, États-Unis) a été évaluée dans un mélange contenant 1 µL d'ADNg (~20 ng/µL), 5 µL de tampon DreamTaq, 1 µL de dNTPs (10 mM), 2,5 µL d'amorce fw (10 µM), 2,5 µL d'amorce rv (10 µM), 37,8 µL H₂O, 0,25 µL de DreamTaq DNA polymérase (1,25 unités). La polymérase contenue dans le GoTaq® Long PCR Master Mix (Promega, Madison, États-Unis) a également été évaluée en ajoutant 25 µL du mix à un mélange contenant 1 µL d'ADNg (~20 ng/µL), 2,5 µL d'amorce fw (10 µM), 2,5 µL d'amorce rv (10 µM), et 19 µL H₂O.

Pour les 3 polymérases considérées, les conditions de PCR suivantes ont été appliquées : une dénaturation initiale de 5 min à 95 °C, suivi de 30 cycles de « dénaturation - hybridation des amorces – élongation » (95 °C, 30 sec – 55 °C, 30 sec – 68 °C, 17 min 30 sec), puis une élongation finale de 10 min à 68 °C.

Le fragment de 17 kb utilisé pour la transformation chez *S. cerevisiae* a été amplifié, dans les mêmes conditions, par la polymérase GoTaq® Long PCR Master Mix en utilisant le couple d'amorces *AstS2* / *AstR3-tag-stop*. Le mélange réalisé pour la PCR était le suivant : 25µL de GoTaq® Long PCR Master Mix, 2,5 µL d'amorce fw (10 µM), 2,5 µL d'amorce rv (10 µM), 19 µL H₂O, et 1 µL d'ADNg de *C. asteris* (~20 ng/µL). Le fragment obtenu a été purifié en utilisant le kit GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Boston, États-Unis).

Un protocole spécifique a été utilisé pour la migration sur gel d'agarose des fragments de grandes tailles amplifiés par PCR. Initialement, 5 µL de marqueur de taille GeneRuler high range DNA ladder (Thermo Fisher Scientific, Boston, États-Unis) était déposé, en parallèle des échantillons, avec une migration à 100 V de 30 min sur gel d'agarose à 0,8 % Tris acétate EDTA. Le protocole modifié pour favoriser la séparation des fragments dans le marqueur de taille implique le dépôt de 1 µL de marqueur de taille GeneRuler high range DNA ladder, dilué dans 5 µL H₂O, pour une migration à 50 V de 2h sur un gel d'agarose à 0,3 % Tris acétate EDTA.

2.5. Analyse de l'adaptabilité des codons chez les levures

L'adaptabilité des codons des souches *S. cerevisiae*, *H. polymorpha*, *P. pastoris* et *Y. lipolytica* vis-à-vis du gène *AstS* a été évaluée par l'intermédiaire de l'outil « graphical codon usage analyser » (<http://gcua.schoedl.de/>). À titre de comparaison,

l'adaptabilité des codons de *S. cerevisiae* vis-à-vis de *pcbAB* (exprimé de manière hétérologue dans une précédente étude [21]) a été analysée selon la même méthode.

2.6. Assemblage des plasmides

2.6.1. Plasmides et amorces utilisés

Les différents plasmides et amorces utilisés au cours de cette étude et décrits dans la suite de cette section sont présentés respectivement dans les **Tableau 1** et **Tableau 2**.

Tableau 1 : Plasmides utilisés ou assemblés au cours de cette étude.

Plasmide	Description du plasmide	Référence / Provenance
pGEM-t-easy	Vecteur de clonage <i>E. coli</i> <i>AmpR, LacZ</i>	Promega
pGEM-Ast5'	<i>Amp, LacZ, Ast5'</i>	Cette étude
pGEM-Ast3'	<i>Amp, LacZ, Ast3'</i>	Cette étude
pGEM-Ast'5'	<i>Amp, LacZ, Ast'5'</i>	Cette étude
pGEM-Ast'3'	<i>Amp, LacZ, Ast'3'</i>	Cette étude
<hr/>		
pESC-ura	Vecteur de clonage <i>S. cerevisiae</i> <i>AmpR, Ura3, pGAL1/GAL10</i>	Stratagene
pESC-ura-sfp	<i>AmpR, Ura3, pGAL1/GAL10, sfp</i>	Cette étude
pESC-ura-sfp-Ast5'	<i>AmpR, Ura3, pGAL1/GAL10, sfp, Ast5'</i>	Cette étude
pESC-ura-sfp-Ast5'-Ast3'	<i>AmpR, Ura3, pGAL1/GAL10, sfp, Ast5', Ast3'</i>	Cette étude
pESC-sfp-AstS.PCR	<i>AmpR, Ura3, pGAL1/GAL10, sfp, astS</i> (obtenu par amplification PCR)	Cette étude
pESC-ura-sfp-AstS.RESTR	<i>AmpR, Ura3, pGAL1/GAL10, sfp, astS</i> (obtenu par restriction de l'ADNg)	Cette étude
<hr/>		
pYX112	Vecteur de clonage <i>S. cerevisiae</i> <i>AmpR, Ura3, pTPI</i>	Ingenius
pYX112-Ast'5'	<i>AmpR, Ura3, pTPI, Ast'5'</i>	Cette étude
pYX112-Ast'5'-Ast'3'	<i>AmpR, Ura3, pTPI, Ast'5', Ast'3'</i>	Cette étude
pYX-AstS.RESTR	<i>AmpR, Ura3, pTPI, astS</i> (obtenu par restriction de l'ADNg)	Cette étude
<hr/>		
RIP191	Vecteur de clonage <i>Y. lipolytica</i> <i>pEYK300A3B, URA3ex</i>	Cette étude
JMP4137	<i>JMP62- pEYK300A3B, YFP, URA3ex</i>	[22]
RIP222	<i>pEYK300A3B, LEU2ex</i> (provenant de JMP114)	Cette étude
JMP114	pBluescript II – cassette avec fragment 1,8 kb <i>LEU2</i>	[23]
RIP191-Ast6kb	<i>pEYK300A3B-URA3ex, Ast6kb Ast785-rv</i>	Cette étude
RIP191-Ast.Tr	<i>pEYK300A3B-URA3ex, Ast.Tr (Ast6kb-Cterm)</i>	Cette étude
RIP222-NpgA	<i>pEYK300A3B, LEU2ex, npgA</i>	Cette étude

Tableau 2 : Séquences des différentes amorces utilisées au cours de cette étude. Les sites de restriction et les tag 6xHIS ajoutés par le biais des amorces sont respectivement indiqués en orange et en vert.

Gène/plasmide associé	Nom de l'amorce	Séquence (5'-3')	Tm (°C)
pGEM-t-easy	<i>M13-fw</i>	gtaa aacgacg gccag	50
	<i>M13-rv</i>	cagg aaacagc tatgac	46
pYX112	<i>pyx-verif-fw</i>	agggcagc ataatttag gagt	53
	<i>Pyx-verif-rv</i>	gctcgagat atcatgc tagtc	55
pESC-ura	<i>Gall0-ProF</i>	ggtggt aatgcc atgtaatatg	51
	<i>pESC-ADH-verif</i>	ccgaca acctgattg gagact	57
RIP191 / RIP222	<i>Zeta1</i>	tgtgtcggacc gcgttc agggtggaacag	68
	<i>Zeta2</i>	actgtcggact ttgtgag agaggtaacgccg	66
	<i>tLip2-rv</i>	gtcttagagga acgc atatacag	53
<i>sfp</i>	<i>SallI-sfp-fw</i>	cg c-gtcgac -tgcagacggaggtctagaatg	67
	<i>SacII-sfp-rv</i>	g cg-ccg cgg-tcattattatcatgattat	60
	<i>qp-sfp-rv</i>	gtgaatcaa acgc accaatg	53
<i>NpgA</i>	<i>HindIII-NpgA-fw</i>	cg c-aagctt -gacatggtgcaagacacatcaa	65
	<i>AvrII-NpgA-rv</i>	cg c-cctagg -cggttaggataggcaattacaca	55
<i>AstS</i>	<i>AstS1</i>	cggacagaatgatacagagca	54
	<i>AstS2</i>	acatttctccttgacctccac	54
	<i>Ast5-bgn-fw</i>	tacacatccgacatgatgaa	53
	<i>AstR5-restr</i>	gcttcgagacttgaggtgtaatc	55
	<i>AstR5-restr-HpaI-rv</i>	cg -g ttaac-gcttcgagacttgaggtgtaatc	63
	<i>AstF3</i>	cggatctcatg tcagg actctccga	61
	<i>AstF3 -HpaI</i>	cg -g ttaac-cggatctcatg tcagg actctccga	61
	<i>AstR3-tag-stop</i>	aatccttcg tcg cagagcaccaaca	61
	<i>AstR3-tag-HIS-stop</i>	cgctcatca -gtggtggtggtggtg g-gtccctcgtcagagcaccaaca	61
	<i>3'-rev17</i>	caactactcaagtgaccgaa	54
<i>Ast785-rv</i>	aagcatagagggacaatgtgg	54	
<i>BamHI-Ast6kb-fw</i>	cg c-gg atcc-gacatgatgaaatcgacctgtact	51	
<i>Ast6kb-sfiI-sfiI-AvrII</i>	g cg-cctagg -ata-g gccg gcttggcc-aat-g gcc catg ccg gcc-aacgctgcctttctcaaa	52	
<i>SfiA-Cterm-fw</i>	cg-g ccg catg ggc cc-gaaccactgggtctgcagtatgctttcttaga	63	
<i>SfiB-Cterm-rv</i>	cg -g ccg gcttggc cc-tcatca gtggtggtggtggtg gtgctctcg	68	

2.6.2. Méthodes communes à l'assemblage de l'ensemble des plasmides

2.6.2.1. Amplification PCR

Les fragments amplifiés par PCR (« polymerase chain reaction »), destinés à être clonés au sein d'un plasmide, ont été obtenus en utilisant l'ADN polymérase à haut taux de fidélité Q5[®] High-Fidelity DNA Polymerase (New England Biolabs - NEB, Beverly, États-Unis). Le mélange et les conditions réactionnelles utilisés pour réaliser les amplifications sont présentés respectivement dans le **Tableau 3A** et **Tableau 3B**.

Tableau 3 : Conditions générales des amplifications PCR réalisées avec l'ADN polymérase Q5[®] High-Fidelity DNA Polymerase en termes de composition du mélange (A) et de réaction réalisée (B).

A		B			
Composant	Volume (µL)	Étape	Nombre de cycles	Température (°C)	Temps (sec)
Q5 Buffer (5X)	10	Dénaturation initiale	1	98	30
dNTPs (10 mM)	1	Dénaturation		98	15
Amorce fw (10 µM)	2,5	Hybridation des amorces	30	50-68*	30
Amorce rv (10 µM)	2,5	Extension		72	30**
Q5 [®] HF DNA Pol	0,5	Extension finale	1	72	2
ADN matrice	Variable (~50 ng)	Stockage	1	4	∞
H ₂ O	qsp 50 µL				

*T°C d'hybridation (T_m) spécifique au couple d'amorces

**Durée de 30 sec par kilobase de séquence à amplifiée

Les amplifications PCR réalisées pour vérifier l'insertion d'un fragment ou confirmer l'obtention de transformant positif chez *E. coli* ont été effectuées via la Taq DNA polymérase (NEB, Beverly, États-Unis). Le mélange et les conditions réactionnelles utilisés pour réaliser les amplifications à l'aide de cette polymérase sont présentés respectivement dans le **Tableau 4A** et **Tableau 4B**.

Tableau 4 : Conditions générales des amplifications PCR réalisées avec la Taq DNA polymérase en termes de composition du mélange (A) et de réaction réalisée (B).

A		B			
Composant	Volume (µL)	Étape	Nombre de cycles	Température (°C)	Temps (sec)
Standard taq Buffer (10X)	2,5	Dénaturation initiale	1	95	30
dNTPs (10 mM)	0,5	Dénaturation		95	30
Amorce fw (10 µM)	0,5	Hybridation des amorces	30	50-68*	60
Amorce rv (10 µM)	0,5	Extension		68	60**
Taq DNA Pol.	0,125	Extension finale	1	68	5
ADN matrice	Variable ^a	Stockage	1	4	∞
H ₂ O	qsp 25 µL	*T°C d'hybridation (Tm) spécifique au couple d'amorces			
^a colonie ou ~50 ng de plasmide		**Durée de 60 sec par kilobase de séquence à amplifier			

2.6.2.2. Purification des fragments PCR

Les fragments d'ADN obtenus par amplification PCR dans le cadre de cette étude ont été visualisés après migration sur gel d'agarose 1 % (Migration 100 V, 20-30 min, Electrophorèse RunOne, VWR International, Darmstadt, Germany). Le colorant de charge Blue / Orange Loading Dye (Promega, Madison, États-Unis) a été préalablement ajouté à l'échantillon déposé (17 % du volume total). Le marqueur de taille Smartladder (Eurogentech, Seraing, Belgique) a été utilisé pour estimer la taille des fragments amplifiés (en pb). La visualisation du profil de migration et l'acquisition des images ont été réalisées via le trans-illuminateur GenoSmart (VWR International, Darmstadt, Germany).

Les fragments PCR obtenus ont été purifiés après migration sur gel en utilisant le kit Monarch[®] DNA Gel Extraction (NEB, Beverly, États-Unis). Les concentrations en ADN obtenus ont été déterminées par un dosage fluorimétrique via Qubit 3.0 (Invitrogen, Carlsband, États-Unis).

2.6.2.3. Déphosphorylation et ligation au sein des plasmides

Préalablement à l'étape de ligation, les plasmides linéarisés ont été déphosphorylés selon les instructions du kit Rapid DNA Dephos & Ligation Kit (Sigma-Aldrich, St Louis, États-Unis). La ligation entre le fragment PCR et le plasmide linéarisé déphosphorylé a été réalisée selon les instructions du même kit, à 20 °C pendant 12 h. Les ligations ont été réalisées avec un ratio molaire plasmide:insert de 1:3. La quantité d'insert (Q_{insert}) à ajouter dans le produit de ligation (en ng) a été calculée comme suit

(avec L_{insert} : longueur de l'insert en kb ; C_{plasmide} : concentration du plasmide en ng/ μL ; L_{plasmide} : longueur du plasmide en kb) :

$$Q_{\text{insert}} = 3 \times \frac{(L_{\text{insert}} \times C_{\text{plasmide}})}{L_{\text{plasmide}}}$$

Dans le cas d'un clonage au sein de pGEM-t-easy (Promega, Madison, États-Unis), le protocole de clonage spécifique à ce plasmide a été suivi. La ligation au sein de ce plasmide se fait par la méthode TA cloning, nécessitant l'ajout d'une base A aux extrémités du fragment. Cette opération a été réalisée en incubant le fragment PCR purifié 20 min à 72 °C dans un mélange contenant : 0,3 μL Taq DNA Polymérase (1,5 unités), 1 μL dATP (10 μM), 5 μL MgCl_2 buffer, qsp 50 μL H_2O (NEB, Beverly, États-Unis).

2.6.2.4. Transformation chez *E. coli* et sélection des transformants

La souche *E. coli* DH5 α thermo-compétente, a été utilisée pour réaliser les différentes transformations. Le produit de ligation (4 μL) a été ajouté dans 100 μL de suspension cellulaire. La méthode classiquement utilisée pour la transformation des cellules comprend une première étape d'incubation des cellules sur glace (30 min), suivi d'un choc thermique à 42 °C (1 min), puis d'une nouvelle incubation sur glace (2 min). Un volume de 900 μL de milieu S.O.C (Invitrogen, Carlsband, États-Unis) a été ajouté aux 100 μL de la suspension cellulaire, suivi d'une incubation de 60 min à 37 °C sous agitation légère à 50 rpm, afin de permettre une régénération des cellules.

Les cellules transformées sont ensuite étalées (100 μL par boîte) sur milieu LB supplémenté en antibiotique afin de sélectionner uniquement les transformants positifs (ampicilline ou kanamycine selon le plasmide considéré). Dans le cas d'une transformation impliquant le plasmide pGEM-t-easy, 50 μL d'IPTG (à 100 mM) et 100 μL de X-Gal (à 20 mg/mL), ont été préalablement étalés sur le milieu LB-ampicilline gélosé afin d'éliminer les faux positifs par une sélection blanc / bleu (complémentation par *lacZ*). Les transformants positifs ont ensuite été confirmés par PCR sur colonie en utilisant l'une des amorces spécifiques au plasmide considéré et une amorce spécifique à la séquence insérée. Pour cela une partie de la colonie est prélevée et mise en suspension dans 25 μL d'un mix PCR. Les cellules sont lysées au début de la PCR à une température de 95 °C pendant 10 min.

2.6.2.5. Extraction et purification des plasmides

L'extraction des plasmides à partir d'*E. coli* a été réalisée via le kit Wizard Plus Minipreps DNA Purification System (Promega, Madison, États-Unis). Les concentrations en plasmide obtenues ont été ensuite déterminées par un dosage fluorimétrique via Qubit 3.0 (Invitrogen, Carlsband, États-Unis).

2.6.3. Différentes étapes nécessaires à l'obtention des plasmides

2.6.3.1. Destinés à une expression chez *S. cerevisiae*

2.6.3.1.1. Plasmide pESC-ura-sfp-Ast5'-Ast3'

Le plasmide pESC-ura-sfp-Ast5'-Ast3' a été obtenu par le clonage successif des fragments *sfp*, *Ast5'* et *Ast3'*. Dans un premier temps, le fragment *sfp* a été amplifié par PCR à partir de l'ADN total de *B. subtilis* BBG111 [19], en utilisant le couple d'amorces *Sall-sfp-fw* / *SacII-sfp-rv* (température d'hybridation [T_m] : 60 °C, temps d'élongation [T_{élong}] : 45 sec). Le fragment de 815 pb obtenu a été purifié sur gel, puis restreint à 37 °C durant 60 min par les enzymes de restriction *Sall* HF[®] et *SacII* (NEB, Beverly, États-Unis). Cette restriction a été réalisée en ajoutant un volume de 40 µL de fragment PCR (~50 ng/µL) au mélange contenant 5 µL de tampon CutSmart[®], 0,5 µL de *Sall* HF[®] (10 unités), 0,5 µL de *SacII* (10 unités), et 4 µL H₂O. Le plasmide pESC-ura a été linéarisé en utilisant les mêmes enzymes de restriction dans le mélange suivant : 20 µL de plasmide (~150 ng/µL), 5 µL de tampon CutSmart[®], 0,5 µL de *Sall* HF[®] (10 unités), 0,5 µL de *SacII* (10 unités), et 24 µL H₂O. Les deux fragments ont été purifiés sur gel, puis mis en contact en présence d'une ligase (ratio plasmide 1:3 insert) durant 12 h. Le plasmide pESC-ura-sfp obtenu a été transformé chez *E. coli* DH5α. La présence de l'insert a été confirmée par l'amplification d'un fragment de 1023 pb en utilisant le couple d'amorces *pESC-ADH-verif* / *qPsfp-rv*.

Le fragment de 1373 pb correspondant à la partie 5' du gène *astS* a été amplifié à partir de l'ADN génomique de *C. asteris* en utilisant le couple d'amorces *Ast5-bgn-fw* / *Ast5-restr-rv* (T_m : 54 °C, T_{élong} : 75 sec). Ce fragment a été purifié et inséré par TA-cloning dans le plasmide pGEM-t-easy, générant le plasmide pGEM-Ast5'. Le fragment Ast5' (1393 pb) a ensuite été obtenu par restriction de 60 min à 37 °C via l'enzyme *EcoRI* HF[®] (NEB, Beverly, États-Unis). Cette restriction a été réalisée en ajoutant un volume de 20 µL de pGEM-Ast5' (~150 ng/µL) au mélange contenant 5 µL de tampon CutSmart[®], 0,5 µL de *EcoRI* HF[®] (10 unités), et 23,5 µL H₂O. Le plasmide pESC-ura-sfp a été linéarisé en utilisant la même enzyme de restriction dans les mêmes conditions. Le plasmide linéarisé et le fragment de 1393 pb ont été purifiés sur gel, puis mis en contact en présence d'une ligase. Le plasmide pESC-ura-sfp-Ast5' obtenu a été transformé chez *E. coli* DH5α. La présence de l'insert dans le sens souhaité a été confirmée par l'amplification d'un fragment de 885 pb en utilisant le couple d'amorces *Gal10-ProF* / *Ast785-rv*.

Le fragment de 648 pb correspondant à la partie 3' du gène *astS* a été amplifié à partir de l'ADN génomique de *C. asteris* en utilisant le couple d'amorces *AstF3* / *AstR3-tag-stop* (T_m : 60 °C, T_{élong} : 40 sec). Ce fragment a été purifié et inséré par TA-cloning dans le plasmide pGEM-t-easy, générant le plasmide pGEM-Ast3'. Le fragment Ast3' (684 pb) a ensuite été libéré par restriction de 60 min à 37 °C via l'enzyme *NotI* HF[®] (NEB, Beverly, États-Unis). Cette restriction a été réalisée en ajoutant un volume de 20 µL de pGEM-Ast3' (~150 ng/µL) au mélange contenant 5 µL de tampon CutSmart[®], 0,5 µL de *NotI* HF[®] (10 unités), et 23,5 µL H₂O. Le plasmide pESC-ura-sfp-Ast5' a été linéarisé en utilisant la même enzyme de restriction dans les mêmes conditions. Le plasmide linéarisé et le fragment de 684 pb

ont été purifiés sur gel, puis mis en contact en présence d'une ligase. Le plasmide pESC-ura-sfp-Ast5'-Ast3' obtenu a été transformé chez *E. coli* DH5 α . La présence de l'insert dans le sens souhaité a été confirmée par l'amplification d'un fragment de 774 pb en utilisant le couple d'amorces *pESC-ADH-verif / AstF3*.

2.6.3.1.2. Plasmide pYX112-Ast'5'-Ast'3'

Le plasmide pYX112-Ast'5'-Ast'3' a été obtenu en clonant successivement les fragments Ast'5' et Ast'3'. Le fragment de 1381 pb correspondant à la partie 5' du gène *astS* a été amplifié à partir du plasmide pESC-ura-sfp-Ast5'-Ast3' en utilisant le couple d'amorces *Ast5-bgn-fw / AstR5-restr-HpaI* (T_m : 54 °C, T_{élong} : 75 sec). Ce fragment a été purifié et inséré par TA-cloning dans le plasmide pGEM-t-easy, générant le plasmide pGEM-Ast'5'. Le fragment Ast'5' (1401 pb) a ensuite été libéré par restriction de 60 min à 37 °C via l'enzyme *EcoRI* HF[®]. Cette restriction a été réalisée en ajoutant un volume de 20 μ L de pGEM-Ast'5' (~150 ng/ μ L) au mélange contenant 5 μ L de tampon CutSmart[®], 0,5 μ L de *EcoRI* HF[®] (10 unités), et 23,5 μ L H₂O. Le plasmide pYX112 a été linéarisé en utilisant la même enzyme de restriction dans les mêmes conditions. Le plasmide linéarisé et le fragment de 1401 pb ont été purifiés sur gel, puis ligués. Le plasmide pYX112-Ast'5' obtenu a été transformé chez *E. coli* DH5 α . La présence de l'insert inséré à la suite du promoteur, a été confirmée par l'amplification d'un fragment de 970 pb en utilisant le couple d'amorces *pyx-verif-fw / Ast785-rv*.

Le fragment de 681 pb correspondant à la partie 3' du gène *astS* a été amplifié à partir du plasmide pESC-ura-sfp-Ast5'-Ast3' en utilisant le couple d'amorces *AstF3-HpaI / AstR3-tag-HIS-stop* (T_m : 60 °C, T_{élong} : 40 sec). Ce fragment a été purifié et inséré par TA-cloning dans le plasmide pGEM-t-easy, générant le plasmide pGEM-Ast3'. Le fragment Ast3' (779 pb) a ensuite été libéré par restriction de 60 min à 37 °C en utilisant les enzymes *ApaI* et *MluI* HF[®] (NEB, Beverly, États-Unis). Cette restriction a été réalisée en ajoutant un volume de 20 μ L de pGEM-Ast'3' (~150 ng/ μ L) au mélange contenant 5 μ L de tampon CutSmart[®], 0,20 μ L de *ApaI* (10 unités), 0,5 μ L de *MluI* HF[®] (10 unités), et 22,5 μ L H₂O. Le plasmide pYX-112-Ast'5' a été linéarisé avec les mêmes enzymes dans les mêmes conditions. Le plasmide linéarisé et le fragment de 779 pb ont été purifiés sur gel, puis mis en contact en présence d'une ligase. Le plasmide pYX112-Ast'5'-Ast'3' obtenu a été transformé chez *E. coli* DH5 α . La présence de l'insert dans le sens souhaité a été confirmée par l'amplification d'un fragment de 790 pb en utilisant le couple d'amorces *pyx-verif-rv / AstF3-HpaI*.

2.6.3.2. Destinés à une intégration chez *Y. lipolytica*

2.6.3.2.1. Plasmide RIP191-Ast6kb-Cterm

Le plasmide RIP191 est un plasmide dérivé du plasmide JMP4137 (JMP62-pEYK300A3B-YFP-URA3ex) [22] dont la séquence correspondante à *yfp* a été excisée. Le plasmide a été linéarisé par restriction *BamHI* HF[®] et *AvrII* (NEB, États-Unis) à 37 °C pendant 60 min. Cette restriction a été réalisée en ajoutant un volume

de 20 μL de RIP191 ($\sim 150 \text{ ng}/\mu\text{L}$) au mélange contenant 5 μL de tampon CutSmart[®], 0,5 μL de *Bam*HI HF[®] (10 unités), 2 μL de *Avr*II (10 unités), et 21,5 μL H₂O.

Le fragment du gène *astS* codant pour les deux premiers modules de la NRPS a été amplifié par PCR à partir du plasmide pYX112.AstS.RESTR en utilisant le couple d'amorces *Bam*HI-*Ast6kb-fw* / *Ast6kb-sfiA-sfiB-AvrII-rv* (T_m : 65 °C, T_{élong} : 240 sec). Le fragment de 6012 pb obtenu a été purifié sur gel, puis restreint à 37 °C durant 60 min par les enzymes de restriction *Bam*HI HF[®] et *Avr*II. Cette restriction a été réalisée en ajoutant un volume de 40 μL de fragment PCR ($\sim 50 \text{ ng}/\mu\text{L}$) au mélange contenant 5 μL de tampon CutSmart[®], 0,5 μL de *Bam*HI HF[®] (10 unités), 2 μL de *Avr*II (10 unités), et 1,5 μL H₂O. Le fragment PCR restreint et le plasmide RIP191 linéarisé ont été purifiés sur gel, puis mis en contact en présence d'une ligase. Le plasmide RIE191-Ast6kb obtenu a été transformé chez *E. coli* DH5 α . La présence de l'insert a été confirmée en amplifiant sur le plasmide extrait un fragment de 2840 pb avec le couple d'amorces *Zeta1* / *Ast785-rv*.

Le fragment du gène *astS* codant pour le domaine C_{term} de la NRPS a été amplifié par PCR à partir du plasmide pYX112.AstS.RESTR en utilisant le couple d'amorces *SfiA-Cterm-fw* / *SfiB-Cterm-rv* (T_m : 68 °C, T_{élong} : 60 sec). Le fragment de 1229 pb obtenu a été purifié sur gel, puis restreint à 37 °C durant 60 min par l'enzyme *Sfi*I (NEB, Beverly, États-Unis). Cette restriction a été réalisée en ajoutant un volume de 40 μL de fragment PCR ($\sim 50 \text{ ng}/\mu\text{L}$) au mélange contenant 5 μL de tampon CutSmart[®], 0,5 μL de *Sfi*I (10 unités), et 4,5 μL H₂O. Le plasmide RIP191-Ast6kb a été linéarisé en utilisant les mêmes enzymes de restriction dans le mélange suivant : 20 μL de plasmide ($\sim 150 \text{ ng}/\mu\text{L}$), 5 μL de tampon CutSmart[®], 0,5 μL de *Sfi*I (10 unités), et 24,5 μL H₂O. Le fragment PCR restreint et le plasmide RIP191-Ast6kb linéarisé ont été purifiés sur gel, puis mis en contact en présence d'une ligase. Le plasmide RIE191-Ast6kb-Cterm obtenu a été transformé chez *E. coli* DH5 α . La présence de l'insert a été confirmée en amplifiant sur le plasmide extrait un fragment de 808 pb avec le couple d'amorces *tLip2-rv* / *AstF3*.

2.6.3.2.2. Plasmide RIP222-NpgA

En premier lieu, la cassette *LoxP*-URA3-*LoxR* a été excisée du plasmide RIP191 en utilisant l'enzyme de restriction *I-Sce*I (NEB, Beverly, États-Unis) à 37 °C durant 60 min. Cette restriction a été réalisée en ajoutant un volume de 20 μL de RIP191 ($\sim 150 \text{ ng}/\mu\text{L}$) au mélange contenant 5 μL de tampon CutSmart[®], 2 μL de *I-Sce*I (10 unités), et 23 μL H₂O. La séquence excisée a ensuite été remplacée chez RIP191 par la cassette *LoxP*-LEU2-*LoxR* obtenue à partir d'une restriction dans les mêmes conditions du plasmide JMP114 [23], générant ainsi le plasmide RIP222.

Le fragment *npgA* a été amplifié par PCR à partir de l'ADN génomique d'*A. nidulans* (IHEM 22446) en utilisant le couple d'amorces *Hind*III-*NpgA-fw* / *Avr*II-*NpgA-rv* (T_m : 60 °C, T_{élong} : 60 sec). Le fragment de 1059 pb obtenu a été purifié sur gel, puis restreint à 37 °C durant 60 min par les enzymes de restriction *Hind*III HF[®] et *Avr*II (NEB, Beverly, États-Unis). Cette restriction a été réalisée en ajoutant un volume de 40 μL de fragment PCR ($\sim 50 \text{ ng}/\mu\text{L}$) au mélange contenant 5 μL de tampon

CutSmart[®], 0,5 µL de *Hind*III HF[®] (10 unités), 2 µL de *Avr*II (10 unités), et 2,5 µL H₂O. Le plasmide RIP222 a été linéarisé en utilisant les mêmes enzymes de restriction dans le mélange suivant : 20 µL de plasmide (~150 ng/µL), 5 µL de tampon CutSmart[®], 0,5 µL de *Hind*III HF[®] (10 unités), 2 µL de *Avr*II (10 unités), et 22,5 µL H₂O. Les deux fragments ont été purifiés sur gel, puis mis en contact en présence d'une ligase. Le plasmide RIP222-NpgA obtenu a été transformé chez *E. coli* DH5α. La présence de l'insert a été confirmée par l'amplification d'un fragment de 566 pb en utilisant le couple d'amorces *tLip2-rv* / *qP1-npg-fw*.

2.7. Obtention des transformants de levure

2.7.1. Transformation chez *S. cerevisiae*

2.7.1.1. Méthode de clonage (TAR Cloning)

La souche *S. cerevisiae* D5y5 a été transformée avec trois plasmides différents : pESC-sfp-AstS.PCR, pESC-sfp-AstS.RESTR, et pYX112-AstS.RESTR. L'insertion de la séquence *astS* de 17 kb au sein des plasmides linéarisés pESC-sfp-Ast5'-Ast3' et pYX112-Ast5'-Ast3' a été réalisée via la méthode du TAR Cloning [24]. Les plasmides pESC-sfp-Ast5'-Ast3' et pYX112-Ast5'-Ast3' ont été respectivement linéarisés en utilisant les enzymes *Pml*I et *Hpa*I (NEB, Beverly, États-Unis). Dans les deux cas, 50 µL de plasmide (~150 ng/µL) ont été ajoutés dans un mélange contenant 15 µL de tampon CutSmart, 20 unités d'enzyme de restriction (1 µL pour *Pml*I, 4 µL pour *Hpa*I), et une qsp 150 µL H₂O. Une incubation de 90 min à 37 °C a été réalisée, suivie par une purification sur gel (Monarch[®] DNA Gel Extraction, NEB, Beverly, États-Unis). Les plasmides ont ensuite été déphosphorylés en utilisant le kit Rapid DNA Dephos & Ligation Kit.

Pour l'obtention du plasmide pESC-sfp-AstS.PCR, 500 ng du plasmide pESC-sfp-Ast5'-Ast3' linéarisé par *Pml*I, ont été co-transformés avec 500 ng du fragment de 17 052 pb obtenu par amplification PCR (ratio molaire 1:3).

Pour l'obtention du plasmide pESC-sfp-AstS.RESTR, 200 ng du plasmide pESC-sfp-Ast5'-Ast3' linéarisé par *Pml*I, ont été co-transformés avec 120 ng d'ADNg de *C. asteris* restreint par *Eco*RI (à 6 ng/µL). Une seconde condition a été testée avec la co-transformation de 400 ng du plasmide pESC-sfp-Ast5'-Ast3' linéarisé par *Pml*I avec 200 ng d'ADNg de *C. asteris* restreint par *Sgr*DI (à 10 ng/µL).

Le plasmide pYX112-AstS.RESTR a ensuite été obtenu en co-transformant 100 ng de plasmide pYX112-Ast5'-Ast3' linéarisé par *Hpa*I, avec 300 ng d'un fragment de 17 kb restreint à partir du plasmide pESC-sfp-AstS.RESTR. La restriction a été réalisée en ajoutant 25 µL du plasmide pESC-sfp-AstS.RESTR dans un mélange contenant 5 µL de tampon Fast digest, 10 µL H₂O, 5µL de *Pac*I, et 5 µL de *Fsp*AI (Thermo Fisher Scientific, Boston, États-Unis). La restriction a été réalisée à 37 °C pendant 60 min, puis l'enzyme a été inactivée à 65 °C pendant 5 min.

La transformation de *S. cerevisiae* a été réalisée selon le protocole décrit par Gietz et Schiestl (2007) [25]. Les cellules compétentes ont été mises en solution dans une suspension contenant l'ADN à cloner (jusqu'à 24 µL), 260 µL de PEG 3350 (50 %), 36 µL d'AcLi (1 M), 50 µL de DNA carrier préalablement dénaturé à 100 °C pendant

5 min (2 mg/mL) (UltraPure™ Salmon Sperm DNA Solution, Thermo Fisher Scientific, Boston, États-Unis), et une qsp 360 µL H₂O. La suspension est incubée 60 min à 42 °C, puis centrifugée pour éliminer le surnageant (13 000 x g, 30 sec). Les cellules sont ensuite mises en suspension dans 200 µL H₂O, étalées sur un milieu SD Medium-URA gélosé, et incubées 72 h à 30 °C.

2.7.1.2. Sélection des transformants

Les colonies obtenues sur le milieu de sélection SD Medium-URA gélosé après 72 h sont repiquées sur une boîte contenant le même milieu gélosé. Cette opération permet d'éviter que la présence d'ADN résiduel présent sur la gélose du premier milieu (et pas dans les cellules), ne génère des faux positifs. Après 48 h, les transformants se développant ont été analysés par colonie PCR selon les conditions décrites dans le **Tableau 4**, en utilisant le couple d'amorces *pESC-ADH-verif* / *3'-rev-17* (pour *pESC-sfp-AstS.PCR* et *pESC-sfp-AstS.RESTR*), et *pyv-verif-rv* / *3'-rev-17* (pour *pYX112-AstS.RESTR*). Un fragment de 876 pb ou 908 pb a été amplifié uniquement chez les transformants positifs.

2.7.2. Transformation chez *Y. lipolytica*

2.7.2.1. Méthode de clonage (intégration génomique)

Les cassettes « Zeta - LoxP - *ura3* - LoxR - *pEYK300A3B* - *AstS.Tr* - *tLip2* - Zeta » et « Zeta - LoxP - *leu2* - LoxR - *pEYK300A3B* - *NpgA* - *tLip2* - Zeta » ont été obtenues respectivement à partir de *RIP191-AstS.Tr* et *RIP222-NpgA*, par amplification PCR en utilisant le couple d'amorces *Zeta1* / *Zeta2*, selon les conditions présentées dans le **Tableau 3**. Ces deux cassettes ont été insérées successivement dans le génome de *Y. lipolytica* RIY261 selon le protocole décrit par Chen *et al.* (1997) [26]. Les cellules compétentes (50 µL) sont mises en contact avec 20 µg d'ADN entraîneur et 500 ng des cassettes à transformer à 28 °C durant 15 min. La suspension est ensuite supplémentée avec 350 µL de PEG 40 %, puis incubée à 28 °C pendant 60 min. Après un choc thermique de 10 min à 39 °C, 600 µL d'acétate de lithium (LiAc, 0,1 M) sont ajoutés à la solution. Un volume de 200 µL de suspension cellulaire est ensuite étalé sur milieu gélosé YNB (éventuellement supplémenté en leucine).

2.7.2.2. Sélection des transformants

Les transformants ayant intégré la cassette « *AstS.Tr* » ont été sélectionnés sur milieu YNB supplémenté en leucine. Les colonies obtenues ont été analysées par colonie PCR selon les conditions décrites dans le **Tableau 4**, en utilisant le couple d'amorces *tLip2-rv* / *3'-rev-17*. Un fragment de 934 pb a été amplifié uniquement chez les transformants positifs. La cassette « *NpgA* » a ensuite été insérée dans le génome du transformant B3, ayant intégré la première cassette, sur milieu YNB. Les colonies obtenues ont été analysées par colonie PCR selon les conditions décrites dans le **Tableau 4**, en utilisant le couple d'amorces *tLip2-rv* / *HindIII-NpgA-fw*. Un fragment de 1214 pb a été amplifié uniquement chez les transformants positifs. Un transformant positif, nommé B3R2, a été sélectionné.

2.8. Séquençages et analyses

L'ensemble des constructions plasmidiques générées au cours de cette étude ont été séquencées par GATC Biotech AG (Konstanz, Germany).

Les plasmides pESC-sfp-AstS.PCR, pESC-sfp-AstS.RESTR, et pYX112-AstS.RESTR ont préalablement été extraits de *S. cerevisiae* DSY5 par l'intermédiaire du kit Easy Yeast Plasmid Isolation (Takara Bio Company, Otsu, Japan). Les plasmides extraits ont ensuite été transformés chez *E. coli* One Shot® OmniMAX™, afin d'être multipliés, puis extraits en concentration suffisante pour permettre le séquençage.

La position et l'impact des mutations révélées lors du séquençage sur la construction pESC-sfp-AstS.PCR ont été analysées à partir des modèles tridimensionnels d'AstS (présentés **Chapitre 3 – Section 2.4.1**) via l'outil bioinformatique Missense3D [27].

2.9. Évaluation de la transcription (RT-qPCR)

Au cours de cette étude, les niveaux d'expression des gènes hétérologues chez plusieurs transformants ont été mesurés, en comparaison avec la souche équivalente non-transformée (contrôle négatif). La présence du transcrite de *astS* a été évaluée après 16 et 36 h de culture pour les souches DS.Pe.Ec73 (milieu YNB-Gal) et DS.Py.62 (milieu YNB) en comparaison avec la souche DSY5. Les niveaux de transcription de la séquence *astS.Tr* et du gène *npgA*, chez la souche B3R2, ont été évalués après 8, 12, 18, et 24 h de culture en milieu YPD-ery, en comparaison avec la souche RIY261.

Pour chaque échantillon considéré, les ARNs ont été extraits en utilisant le kit NucleoSpin® RNA suivant le protocole d'extraction adapté pour les levures (Macherey-Nagel, Düren, Germany). Les levures prélevées (~10⁷ cellules) ont été mises en suspension dans 350 µL de tampon d'extraction, additionné de 3,5 µL de β-mercaptoéthanol. La lyse des cellules a été réalisée par dismembrateur en présence de billes de verre, avec 5 cycles de 2 min à 2000 rpm, alternés de 2 min où l'échantillon était maintenu sur glace. Les billes de verre et les cellules ont ensuite été éliminées par centrifugation (10 min, 10 000 x g, 4 °C). Le lysat obtenu est ensuite filtré afin de réduire sa turbidité / viscosité, puis 350 µL d'éthanol sont ajoutés. Le lysat est ensuite déposé sur une colonne NucleoSpin® RNA, permettant la fixation des ARNs. Les échantillons fixés sur la membrane de la colonne sont ensuite successivement désalinisés, traités par une DNase, puis lavés à plusieurs reprises. Les ARNs sont ensuite élués dans 60 µL d'H₂O dépourvu de RNase. La concentration des ARNs et la pureté de l'échantillon sont estimées par dosage spectrophotométrique sur 1 µL d'échantillon.

Les ARN extraits ont ensuite été convertis en ADN complémentaire (ADNc) par l'action d'une reverse transcriptase. Cette opération a été réalisée en ajoutant un volume de 10 µL d'ARN extrait (à 10 ng/µL) à un mélange composé de 4 µL de LunaScript RT SuperMix (5X) et 6 µL d'H₂O. La solution obtenue a été incubée 2 min à 25 °C - 10 min à 55 °C - 1 min à 95 °C, afin de générer les ADNc correspondant.

La quantité d'ADNc obtenue est proportionnelle à la quantité d'ARN mise en jeu au départ. Il est donc possible, en quantifiant l'ADNc pour un gène en particulier, d'estimer les taux de transcriptions initiaux. Cette quantification a été réalisée par l'intermédiaire d'une RT-qPCR ou PCR quantitative en temps réel. Cette méthode permet, par l'intermédiaire d'une sonde fluorescente présente dans le milieu réactionnel, de suivre l'amplification d'un fragment d'ADN d'intérêt au cours d'une réaction PCR. Le niveau de fluorescence émis peut être corrélé avec la quantité d'ADNc présent tout au long de l'amplification PCR. Le kit utilisé au cours de cette étude (Luna[®] Universal qPCR Master Mix, NEB, Beverly MA, USA) contient la sonde fluorescente SYBR[®] Green se fixant à l'ADN qui est sous sa forme double brin (durant l'étape d'élongation de la PCR). Le mélange utilisé pour la réaction qPCR est le suivant : 1 µL ADNc - 10 µL Universal qPCR Master Mix – 10 µL amorce sens (0,25 µM) – 10 µL amorce antisens (0,25 µM) – qsp 50 µL H₂O.

L'efficacité, à différents degrés de dilution, des couples d'amorces utilisés pour l'amplification de chaque gène a été préalablement évaluée via une qPCR sur un contrôle positif (gène séquencé isolé au sein d'un plasmide). Les couples d'amorces utilisés pour réaliser les qPCR, en fonction du gène dont l'expression est évaluée, sont présentés dans le **Tableau 5**.

Tableau 5 : Liste et séquences des amorces utilisées dans les réactions qPCR pour l'évaluation de la transcription des différents gènes.

Nom de l'amorce	Sens	Séquence (5'–3')	Séquence associée	Longueur amplicon (pb)
<i>FBA1-fw</i>	Fw	GGTTTGTACGCTGGTGACATCGC	<i>fba1</i>	125
<i>FBA1-rv</i>	Rv	CCGGAACCACCGTGGAAGACCA		
<i>qPAstS1-fw</i>	Fw	CCATCAAGGTGTTGTGAACG	<i>astS</i>	237
<i>qPAstS1-rv</i>	Rv	AGCTGCAACCGGTCTCTAAA		
<i>qPAend1</i>	Fw	TGGCTTTTTATCCGGAATTGAGG	<i>astS.Tr</i>	176
<i>qPAend2</i>	Rv	CAAAAGTTCATCCGCACCTTCA		
<i>qp-npg1-fw</i>	Fw	ACTTACTGGGCGCTCAAAGA	<i>npgA</i>	231
<i>qp-npg1-rv</i>	Rv	TAGGTAATCACCGCCAGAG		

Les niveaux de fluorescence durant la réaction ont été mesurés par l'intermédiaire du thermocycleur StepOnePlus[™] Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). La réaction de PCR débute par une dénaturation initiale de 60 sec à 95 °C, puis 40 cycles de dénaturation (15 sec à 95 °C) – extension (30 sec à 60 °C). La réaction s'achève par un gradient à 60-95 °C d'une durée de 60 sec, permettant de définir les courbes de fusion (melting curve). Cette courbe de fusion est un indicateur de la spécificité des amorces utilisées vis-à-vis de l'échantillon analysé.

À l'issue de la réaction PCR, un cycle seuil (Ct) est déterminé pour chaque échantillon analysé. Ce cycle seuil correspond au nombre de cycles nécessaires pour

que la fluorescence mesurée soit supérieure à celle du bruit de fond. Dans le cadre d'un gène exprimé de manière hétérologue, la présence du transcrit peut être confirmée en comparant les Ct obtenus chez le transformant avec ceux obtenus pour la souche non-transformée (où le gène n'est pas présent). Des études précédentes ont montré que dans le cas où l'on souhaite discriminer deux échantillons, l'un où le transcrit est potentiellement présent et l'autre pas, une valeur de coupure pour le Ct peut être fixée à 30 [28]. La présence du transcrit est confirmée pour un Ct en dessous de 30 cycles, et infirmée au-delà de cette valeur.

Dans le cas du gène *astS* exprimé chez *S. cerevisiae* via deux constructions plasmidiques différentes, les Ct obtenus chez les différents transformants ont été comparés entre eux afin de mesurer les niveaux relatifs d'expression. La quantité d'ARN relative obtenue chez la souche DS.Py.62 a été calculée par rapport à une condition contrôle (fixée comme étant l'expression chez le transformant DS.Pe.Ec73). La méthode utilisée, développée par Livak et Schmittgen (2001) [29], consiste en une double normalisation de la quantité d'ARN dans un échantillon, dans un premier temps vis-à-vis d'une condition contrôle, et dans un second temps par rapport à un gène de référence (ici *Fba1*) dont les niveaux d'expression restent stables dans les conditions de l'expérience [30]. Cette méthode permet de calculer un facteur $2^{-\Delta\Delta Ct}$, où la valeur 2 est reliée à l'efficacité de l'amplification PCR qui double la quantité d'ADNc à chaque cycle (efficacité théorique de 100 %), et la valeur $\Delta\Delta Ct$ est déterminée comme suit :

$$\Delta\Delta Ct = (Ct_{\text{échantillon}} - Ct_{\text{référence}}) - (Ct_{\text{échantillon}} - Ct_{\text{contrôle}})$$

Avec $Ct_{\text{échantillon}}$: Ct obtenu pour le gène d'intérêt dans la condition considérée,
 $Ct_{\text{référence}}$: Ct obtenu pour le gène de référence dans la condition considérée, et
 $Ct_{\text{contrôle}}$: Ct obtenu pour le gène d'intérêt dans la condition contrôle.

2.10. Analyses protéomiques

2.10.1. Préparation des lysats protéiques

Les protéines extraites des souches *S. cerevisiae* DSy5, DS.Pe.Ec73, et DS.Py.62 ont été obtenues à partir de cellules cultivées 36 h sur milieu YNB (supplémenté ou non en galactose). Les protéines extraites des souches de *Y. lipolytica* RIY261, et B3R2 ont été obtenues à partir de cellules cultivées 8 ou 24h sur milieu YPD (supplémenté en erythritol).

Après centrifugation (10 000 x g, 5 min) le milieu de culture est éliminé et un volume équivalent de tampon de lyse est ajouté afin de laver les cellules, puis ce tampon est éliminé de nouveau par centrifugation. La masse du culot de cellules est ensuite mesurée afin d'y ajouter une masse égale de billes de verre (0,10 mm de diamètre). Les cellules et billes de verres sont ensuite mise en suspension dans un volume de tampon de lyse équivalent à 8 fois la masse des cellules.

Deux tampons de lyse différents ont été utilisés pour l'extraction des protéines. Le tampon A est composé de phosphate de sodium (50 mM, pH 7,4) additionné de 5 % de glycérol. Le tampon B, plus adapté à l'extraction de protéines membranaires – ou

associées aux membranes, est composé de Tris HCl (10 mM, pH 7,4), NaCl (500 mM), EDTA (1 mM), Triton x 100 (1 %), NP-40 (1 %), deoxycholate de sodium (0,5 %), SDS (0,1 %), glycérol (5 %), dithiothréitol (1 mM). Les composés utilisés ont été fournis par Sigma-Aldrich (St. Louis, États-Unis). Les tampons utilisés ont été supplémentés avec le cocktail inhibiteur de protéase cOmplete™, EDTA-free (5 % du volume total) (Roche Applied Science, Basel, Suisse).

La lyse a été réalisée en utilisant un dismembrateur avec 5 cycles de 2 min à 2000 rpm, alternés de 2 min où l'échantillon était maintenu sur glace. Les billes de verre et les cellules ont été éliminées par centrifugation (10 min, 10 000 × g, 4 °C), et le tampon contenant les protéines a été prélevé et conservé à 4 °C.

La quantité de protéine obtenue a été déterminée par l'intermédiaire d'un dosage colorimétrique de Bradford (Pierce™ Coomassie Protein Assay Kit, Thermo Fisher Scientific, Boston, États-Unis). Une gamme de dosage de la protéine BSA (Sigma-Aldrich, St. Louis, États-Unis), de 0 à 2 mg/mL, a été réalisée pour déterminer la relation entre la DO mesurée et la concentration en protéine.

Les lysats protéiques trop faiblement concentrés (dosage inférieur à 1,5 mg/mL en diluant 10 x dans le tampon de lyse) ont été concentrés en utilisant l'Integrated SpeedVac System (Thermo Fisher scientific, Asheville, États-Unis) avec rotation à 40 °C (« Medium »), jusqu'à atteindre la concentration désirée.

2.10.2. Gel de protéines SDS-Page

Le profil des lysats protéiques obtenus a été analysé après une migration sur gel SDS-Page 7,5 % Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad Laboratories, Hercules, États-Unis). La solution déposée dans chaque puits (20 µL) est composée de 10 µL d'échantillon protéique (à 2 µg/µL) et 10 µL de tampon de charge (composition pour 100 mL : 10 mL Tris HCl 1,25 M pH 6,8, 30 mL SDS 10 %, 10 mL 2-mercaptoéthanol, 40 mg 2-bromophénol, qsp 100 mL H₂O). Le tampon de migration utilisé est composé de Tris (25 mM), glycine (192 µM), et SDS (1 %). Afin de dénaturer les protéines avant le dépôt, la solution est incubée 2 min à 100°C.

La migration est effectuée dans l'unité d'électrophorèse de Bio-rad Laboratories (Hercules, États-Unis) avec une première étape de migration de 20 min à 15 mA, suivie par une seconde étape de 100 min à 25 mA. À l'issue de la migration les protéines présentes dans le gel sont colorées au bleu de Coomassie (composition : 0.5 % bleu de Coomassie G-250, 50 % méthanol, 10 % acide acétique) pendant 60 min. Le colorant fixé au gel de manière non-spécifique est ensuite retiré en incubant le gel dans une solution de décoloration (composition : 40 % méthanol, 10 % acide acétique) pendant 180 min. Le traitement et l'acquisition des images ont été réalisés par l'intermédiaire du visionneur de gel Fusion FX Chemiluminescence Imaging (Vilber, Lourmat, Germany) en mode Coomassie blue (détection automatique).

La taille des protéines observées après migration a été estimée par rapport soit au marqueur de taille Novex™ Sharp Pre-stained Protein Standard (Invitrogen, Carlsbad, États-Unis), soit aux marqueurs de poids moléculaire thyroglobuline (669 kDa),

ferritine (440 kDa), catalase (232 kDa), et lactate déshydrogénase (140 kDa) contenus dans le kit Electrophoresis Calibration Kit (Fine chemical AB. Uppqala, Suède).

2.10.3. Purification des protéines sur colonne

Les lysats protéiques obtenus à partir des cellules de *S. cerevisiae* DS.Py.62 ont également été appliqués sur une colonne HisTrap™ FF 1 mL (GE Healthcare Bio-Sciences AB, New York, États-Unis). L'objectif étant de purifier uniquement la protéine AstS couplée à un tag 6xHIS. Les différentes phases appliquées successivement sur la colonne sont les suivantes : 5 mL H₂O, 5 mL tampon de fixation, 40 mL lysat protéique (0,5 mg/mL), 15 mL tampon de fixation, 20 mL tampon d'éluion, 5 mL de tampon de fixation, 5 mL d'éthanol 20 % (solution de stockage). Le tampon de fixation est composé de Na₃PO₄ (20 mM), NaCl (500 mM), et imidazole (20 mM). Le tampon d'éluion est composé de Na₃PO₄ (20 mM), NaCl (50 mM), et imidazole (500 mM).

La fraction protéique contenue dans le tampon d'éluion a été dialysée à l'intérieur de boudins de dialyse au seuil de coupure de 3,5 kDa (CelluSep® T1 Tubings Nominal Filter, Triple Red Ltd. Oxon, Angleterre). La fraction dialysée a ensuite été concentrée 100 x (jusqu'à 200 µL) en utilisant l'Integrated SpeedVac System avec rotation à 40 °C, puis analysée par SDS Page selon les conditions décrites précédemment.

2.10.4. Western Blot

Un western blot a été réalisé sur des gels SDS page (avant l'étape de coloration) après migration des lysats protéiques extraits des souches de *S. cerevisiae* DS.Pe.Ec73 et DS.Py.62, et de *Y. lipolytica* B3R2. Le transfert des protéines du gel SDS vers la membrane de nitrocellulose (Hybond ECL, Amersham, Buckingham, Angleterre) a été réalisé via la cuve d'électrophorèse Wet / Tank Blotting system de Bio-rad (Laboratories, Hercules, États-Unis) à 40 mV durant 12 h à 4 °C, en présence de tampon de transfert (composition : Tris 24,8 mM, glycine 190,5 mM, 20 % méthanol). La membrane est ensuite incubée 60 min dans une solution de blotting à 4 % (Blotting-Grade Blocker, Bio-Rad Laboratories, Hercules, États-Unis).

Dans le cas des lysats protéiques obtenus à partir des souches DS.Pe.Ec73 (tag FLAG), la membrane a été incubée 3 h en présence de l'anticorps primaire DYKDDDDK Tag Monoclonal Antibody FG4R IgG2b (Invitrogen, Carlsbad, États-Unis) dilué au 1 : 500 dans la solution de blotting. Après avoir retiré la solution de blotting contenant l'anticorps primaire et réalisé un lavage au TBS-Tween (50 mM Tris, 150 mM NaCl, 0.1% Tween 20), la membrane est incubée 1 h avec l'anticorps secondaire goat anti-mouse IgG – F (ab')₂–HRP (SantaCruz Biotechnology, Santa Cruz, États-Unis) dilué au 1 : 5000. La solution DYDDDDK (FLAG®) positive control lysate (Rockland, Limerick, Irlande), déposée dans l'un des puits du SDS Page (2 µL), constitue un contrôle positif du Western blot.

Dans le cas des lysats protéiques obtenus à partir des souches DS.Py.62 et B3R2 (6xHIS tag), la membrane a été incubée 6 h en présence de l'anticorps 6x-His Tag Monoclonal Antibody (HIS.H8) – HRP (Thermo Fisher, Waltham, États-Unis) dilué

au 1 : 500. Le marqueur de poids moléculaire composé de protéines portant un tag 6xHIS (BenchMark™ His-tagged Protein Standard, Invitrogen, Carlsbad, États-Unis) a également été utilisé en tant que contrôle positif du Western blot.

Après 3 lavages successifs de la membrane via du TBS-Tween, la révélation a été réalisée en utilisant le kit de révélation SuperSignal™ West Pico PLUS (Thermo Scientific, Rockford, États-Unis) réagissant en présence de HRP. Le traitement et l'acquisition des images ont été réalisés par l'intermédiaire du visionneur de gel Fusion FX Chemiluminescence Imaging en mode chemiluminescence (détection automatique).

2.10.5. Identification des protéines par spectrométrie de masse

Les bandes spécifiques, visualisées entre 160 et 260 kDa, après migration du lysat protéique de B3R2 sur SDS Page ont été prélevées afin d'être analysées par spectrométrie de masse. La bande de gel SDS Page est tout d'abord décolorée dans un mélange H₂O / AcN, puis incubée 45 min à 56 °C dans une solution de DTT (10 mM) permettant la réduction des ponts-disulfure. Après la réduction, une alkylation des cystéines est réalisée en ajoutant une solution d'iodoacétamide (IAA) (100 mM) et NH₄HCO₃ (50 mM), incubé 30 min à l'obscurité. Les solutions de DTT et d'IAA sont ensuite éliminées, et une solution de trypsine (8 ng/μL dans du tampon NH₄HCO₃ à 50 mM) est ajoutée pour une incubation sur glace durant 30 min pour permettre l'inclusion de la trypsine dans le gel (trypsine sequencing grade, Promega, Madison, États-Unis). L'excès de solution est retiré et remplacé par du tampon NH₄HCO₃ à 50 mM. La digestion par la trypsine est réalisée à 37 °C durant 12 h. Le surnageant est ensuite récupéré après centrifugation (10 000 x g, 10 min) et acidifié par addition de 10 % acide formique.

Les peptides des différents échantillons sont ensuite séchés au SpeedVac, et remis en suspension dans 25 μL H₂O / 0.1 % TFA, puis 10 μL sont injectés en RP HPLC-Q / TOF (avec les paramètres suivants : mode de détection- positif, source d'ions – électrospray (ESI), mode de fragmentation – « collision induced dissociation » (CID), mode scan MS – quadrupôle (Q), et mode scan MS / MS – « time of flight » (TOF)).

Les données obtenues ont été comparées avec le profil peptidique théorique de digestion tryptique de la synthétase d'astine déterminé grâce à l'outil PeptideMass (https://web.expasy.org/peptide_mass/). Les analyses ont été réalisées par l'intermédiaire du logiciel PEAKS Studio 9 [31] (avec les paramètres suivants : query type – homology match, modifications et / ou modifications post-traductionnelles (MTPs) autorisées - Carbamidomethylation – 57,02 Da, variable modifications – oxydation - 15.99 Da, fragment ion tolerance - 0.5 Da, maximum # of MTPs – 3, *De novo* score threshold – 15, peptide hit threshold 10logP – 30).

2.11. Synthèse in vitro du dipeptide 2Abu-Pro

Les souches RIY261 et B3R2 ont été cultivées dans 200 mL de milieu YPD (supplémenté en erythritol) pendant 12 h à 30 °C en agitation à 160 rpm. Le surnageant

a été séparé des cellules par centrifugation (10 000 x g, 10 min), et concentré 10 x en utilisant l'Integrated SpeedVac System avec rotation à 40 °C (« Medium »).

Les cellules ont été récupérées et lysées avec des billes de verre par dismembrateur selon le protocole décrit dans la section 2.9.1 de ce chapitre. Les acides nucléiques ont ensuite été précipités, selon le protocole décrit par [32], en ajoutant de la streptomycine jusqu'à une concentration de 1 %, puis en incubant sur glace pendant 20 min. Le surnageant contenant les protéines a ensuite été séparé des acides nucléiques par centrifugation (10 000 x g, 20 min, 4 °C). Le lysat protéique a ensuite été concentré (jusqu'à 500 µL) sur une membrane d'ultrafiltration Amicon® Ultra-15 Centrifugal filter Devices – 30.000 MWCO (Merck Millipore Ltd., Carrigtwohill, Irlande). La fraction du rétentat contenant les protéines de taille supérieure à 30 kDa a été dialysée 1000 fois dans un tampon de 50 mM Tris-HCl (pH 7,5), puis prélevée.

Un test de synthèse *in vitro* d'un dipeptide L-2Abu - L-Pro a été mis en place d'après un protocole adapté de celui décrit par Haynes *et al.* (2013) [33]. Un mélange (de 40 µL) contenant 6 mM d'ATP, 4 mM de MgSO₄, 2 mM de L-2Abu, et 2 mM de L-Pro a été ajouté dans un volume de 60 µL de la fraction protéique concentrée et dialysée. Le mélange a été incubé 24 h à 20 °C, puis appliqué sur une membrane d'ultrafiltration Amicon® Ultra-15 Centrifugal filter Devices – 10.000 MWCO (Merck Millipore Ltd., Carrigtwohill, Irlande) où seule la fraction du perméat contenant les protéines et peptides de taille inférieure à 10 kDa a été conservée. Cette fraction a ensuite été concentrée 10 x en utilisant l'Integrated SpeedVac System avec rotation à 40 °C (« Medium »).

Le protocole mis en œuvre pour obtenir, à partir de culture de RIY261 et B3R2, un surnageant et un produit de synthèse *in vitro* du dipeptide, est décrit sur la **Figure 2**.

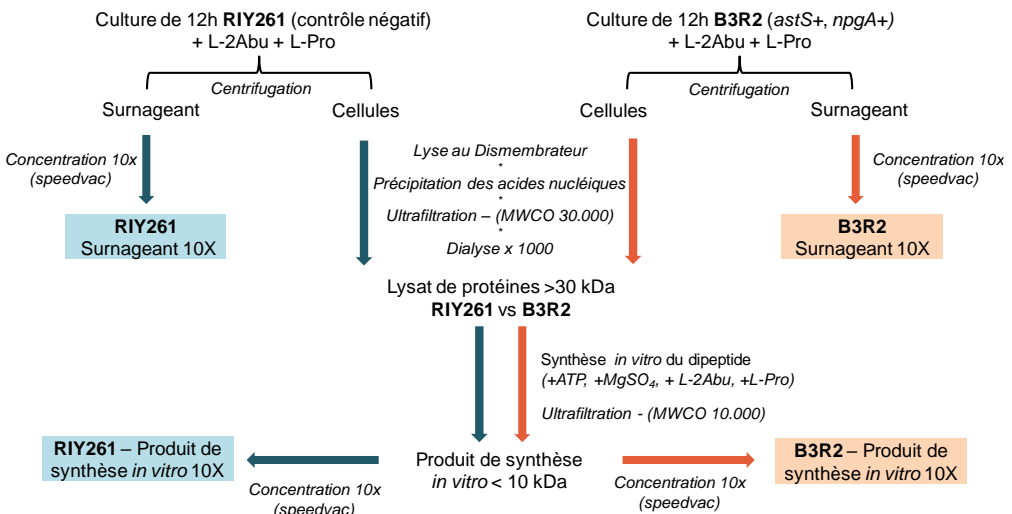


Figure 2 : Protocole appliqué pour générer un surnageant et un produit de synthèse *in vitro* du dipeptide L-2Abu – L-Pro à partir d'une culture de RIY261 et de B3R2.

Le dipeptide L-2Abu – L-Pro a été synthétisé chimiquement par GeneCust (Boynes, France), afin d'être utilisé comme contrôle positif dans la méthode de dosage.

L'analyse de ce composé a été réalisée en injectant 10 µL du standard synthétisé (0,002 M dans H₂O) à l'aide d'un Système HPLC (1290 Series, Agilent Technologies, Waldbronn, Germany) couplé à un QTOF (Q-TOF 6530, Agilent Technologies). Les échantillons ont été analysés à l'aide d'une colonne UPLC (Waters Acquity UPLC BEH C18 1.7 µM (2,1 x 50 mm)) et l'élution a été faite de manière isocratique de 0 à 2 minutes avec un mélange 98 % H₂O / 2 % acétonitrile supplémenté de 0.1 % d'acide formique à un débit de 0,3 mL/min. Les paramètres de la source electrospray (en mode positif) appliqués sont les suivants : température du gaz, 300 °C ; débit de gaz, 8 L/min; pression du nébuliseur, 35 psi ; « Sheath température » 350 °C, « sheath gas flow » 11 L/min ; voltage du fragmenteur : 75 V ; voltage du capillaire : 3500 V ; « Nozzle voltage » : 1000 V. Le « range » de masse allant de m/z 50 à 250 à une vitesse de scan de 1 spectre/seconde.

En premier lieu, cette méthode de dosage a été appliquée pour l'analyse des échantillons « RIY261 - Surnageant 10 X » et « RIY261 – Produit de synthèse *in vitro* », utilisés ici en tant que contrôles négatifs. RIY261 étant la souche non-transformée, la synthétase tronquée n'est pas présente pour permettre l'assemblage d'un dipeptide. Par la suite, afin de confirmer la détectabilité du dipeptide dans les conditions testées, la même analyse a été réalisée en supplémentant le dipeptide synthétisé (à 0,002 M) dans les échantillons constituant les contrôles négatifs. Enfin, les échantillons à tester « B3R2 – Surnageant 10 X » et « B3R2 – Produit de synthèse *in vitro* » ont à leur tour été analysés selon la même méthode.

3. Résultats

3.1. Usage des codons pour l'expression de *astS* chez les levures sélectionnées

L'usage des codons pour l'expression du gène *astS* a été évalué pour les différentes souches de levures considérées pour l'expression hétérologue (*Y. lipolytica*, *S. cerevisiae*, *P. pastoris*, et *H. polymorpha*). L'indice d'adaptabilité du codon (CAI – « Codon Adaptation Index ») a été déterminé chez chaque souche pour chacun des codons composant la NRPS AstS. Les valeurs d'adaptabilité relative sont calculées à partir de la fréquence d'utilisation du codon dans l'organisme considéré [34]. Pour un même acide aminé, le codon présentant la fréquence la plus élevée se verra attribué une adaptativité relative de 100 %, puis l'adaptativité des autres codons est calculée selon leur fréquence relative vis-à-vis du codon le plus fréquent. Plus le nombre de codons à faible adaptabilité est élevé, et plus les taux de traduction peuvent être impactés. Le nombre de codons, sur les 5670 codons de *astS*, pour lesquels l'indice d'adaptabilité est inférieur à 20 % a été évalué pour chaque souche à l'aide de l'outil Graphical Codon Analyser (**Figure 3**).

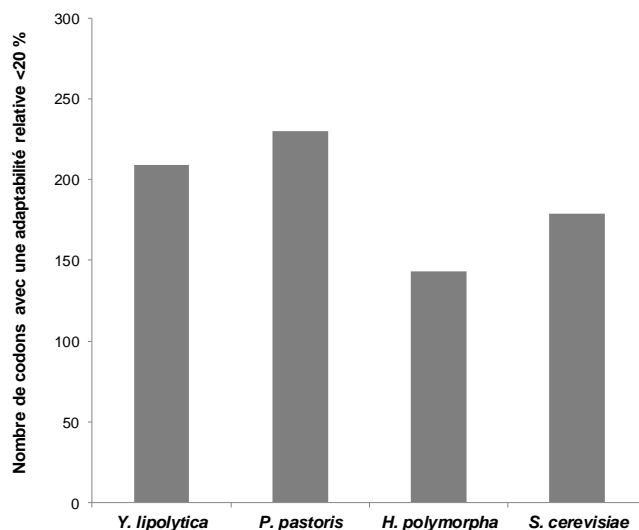


Figure 3 : Nombre de codons du gène *astS* présentant une adaptabilité relative inférieure à 20 % chez *Y. lipolytica*, *P. pastoris*, *H. polymorpha*, et *S. cerevisiae*, déterminé par l'intermédiaire de Graphical Codon Usage Analyser.

Le nombre de codons avec une adaptabilité inférieure à 20 % varie de 143 (pour *H. polymorpha*) à 230 (pour *Pichia pastoris*). Ces valeurs correspondent respectivement à 2,52 % et 4,06 % du nombre de codons total composant le gène *astS*.

Chez *S. cerevisiae* et *Y. lipolytica*, respectivement 179 et 209 codons présentent une adaptabilité inférieure à 20 %. Les codons concernés, leur nombre d'occurrence, et leur adaptabilité spécifique sont présentés dans le **Tableau 6**.

Tableau 6 : Nature et nombre d'occurrence des codons présentant une adaptabilité relative pour l'expression du gène *astS* inférieure à 20 % chez *S. cerevisiae* et *Y. lipolytica*.

Souches	Codons (ADN)	Acide aminé	Nombre d'occurrence	Indice d'adaptabilité relative
<i>S. cerevisiae</i>	CGG	Arg	32	8 %
	CGC		76	13 %
	CGA		71	15 %
<i>Y. lipolytica</i>	CTA	Leu	45	15 %
	TTA		27	5 %
	AGG	Arg	42	12 %
	ATA	Ile	56	8 %
	GTA	Val	39	16 %

Chez les deux souches considérées, un ARNt correspondant est exprimé pour chacun des codons présents dans la séquence *astS*. L'absence d'un ARNt pour l'un des codons (adaptabilité relative de 0 %) empêcherait toute traduction de l'ARNm de *astS*. Chez *S. cerevisiae*, les codons présentant un faible indice d'adaptabilité sont tous associés à l'acide aminé arginine. Afin de déterminer si les taux observés peuvent être considérés comme un facteur limitant à l'étape de traduction, la même analyse d'usage des codons a été réalisée chez *S. cerevisiae* pour la synthétase ACVS. Cette synthétase a déjà été produite de manière hétérologue sans que l'étape de traduction ne soit problématique [21]. Sur les 3777 codons composant la séquence *pcbAB* (codant pour ACVS) chez *Penicillium chrysogenum*, 145 présentent un indice d'adaptativité relative inférieur à 20 % chez *S. cerevisiae*, ce qui représente 3,6 % du nombre de codons total. Dans le cas de la synthétase d'astine, ce chiffre s'élève à 3,13 % (179 / 5670 codons). En comparant avec cette précédente étude, la proportion de codons dits « rare » (à faible indice d'adaptabilité) pour la séquence *astS* ne semble pas constituer un élément capable de faire obstacle à la traduction par les ribosomes.

La proportion de codons rares n'est pas le seul élément à prendre en compte pour évaluer la compatibilité d'usage des codons, puisque la nature de ces codons, et leur position au sein de l'ARNm jouent également un rôle déterminant dans le processus de traduction. Dans le cadre d'une optimisation des codons visant à améliorer les rendements de traduction et de production, l'ensemble de ces éléments serait pris en compte. Néanmoins, l'optimisation d'une séquence en termes d'usage des codons reste un processus couteux et sans garantie de succès. L'approche développée dans cette étude se concentre uniquement sur le transfert de la séquence nucléotidique native de *astS* afin d'évaluer, dans un premier temps, la capacité d'un modèle levure à intégrer et à exploiter la voie non-ribosomique relative à l'astine.

3.2. Extraction de l'ADN de *C. asteris* et isolement du gène codant pour la NRPS AstS

3.2.1. Mise au point d'une méthode d'extraction de l'ADN génomique de *C. asteris*

Cyanodermella asteris est une espèce qui a été isolée et décrite très récemment [35]. À l'heure actuelle, il n'existe pas dans la littérature de protocole spécifique pour l'extraction de son ADN génomique (ADNg). Par conséquent, l'efficacité de trois kits d'extraction distincts ainsi que l'impact de différents pré-traitements de lyse cellulaire ont été évalués. Les pré-traitements considérés sont des procédés de lyses mécaniques induites soit par broyage en présence d'azote liquide, soit par utilisation du dismembrateur avec ajout de billes de verres (lyse mécanique), ou soit par une combinaison de ces deux techniques. Les résultats obtenus sont présentés sur la **Figure 4**.

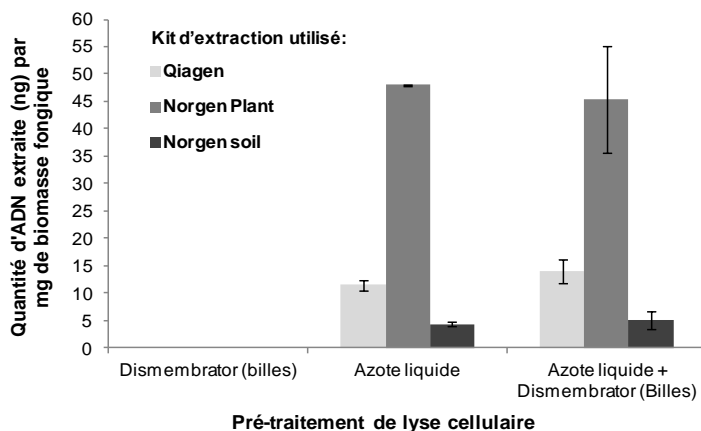


Figure 4 : Impact de différents pré-traitements de lyse cellulaire sur la quantité d'ADN génomique de *C. asteris* obtenu après utilisation des kits Qiagen, Norgen Plant, ou Norgen soil. Chaque extraction a été réalisée en duplicata et les quantités d'ADN sont exprimées en ng par mg de biomasse fongique utilisée lors de l'extraction.

Pour les 3 kits considérés, l'utilisation seule du dismembrateur avec ajout de billes de verre n'a pas permis d'obtenir de l'ADN génomique. Ce procédé de lyse semble inadapté pour les cellules de *C. asteris*. À l'inverse, l'ADNg du champignon a pu être obtenu en broyant les cellules dans de l'azote liquide. L'utilisation du dismembrateur à la suite du broyage en présence d'azote liquide n'a pas permis d'augmenter significativement la quantité d'ADNg extraite.

Après la lyse à l'azote liquide (avec ou sans étape supplémentaire de lyse au dismembrateur), le kit Norgen Plant a permis d'obtenir des rendements en ADN génomique (~ 48 ng/mg de biomasse) plus importants que pour les deux autres kits. L'ADN génomique de *C. asteris* exploité dans la suite de l'étude a été systématiquement obtenu par lyse cellulaire à l'azote liquide suivi d'une extraction via le kit Norgen Plant.

3.2.2. Obtention de la séquence *astS*

Afin de mettre en place une production hétérologue d'astine (ou d'un dérivé d'astine) chez la levure, il est indispensable de transférer la séquence *astS* codant pour la NRPS. La stratégie développée consiste à isoler cette séquence au sein d'un plasmide de clonage afin de pouvoir par la suite facilement la transférer dans d'autres plasmides d'expression afin de l'exprimer sous contrôle d'un promoteur différent ou chez un autre organisme. Étant donné la taille importante de la séquence codante (17 kb), l'utilisation de techniques de biologie moléculaire adaptées est requise. L'approche retenue consiste à mettre à profit la capacité de *S. cerevisiae* à effectuer le TAR cloning avec des rendements très élevés [24]. Le TAR cloning permet d'insérer par recombinaison homologue un fragment d'ADN de grande taille au sein d'un plasmide où les fragments 5' et 3' homologues de ce gène ont préalablement été clonés séparément.

Les rendements de transformation obtenus à la suite du TAR cloning sont influencés par la taille du gène à intégrer, la longueur totale des fragments mis en jeu lors de la transformation, ainsi que leur degré de pureté et leur concentration. Deux approches sont possibles pour obtenir des fragments contenant le gène d'intérêt à co-transformer avec le plasmide linéarisé : par restriction de l'ADN génomique de l'organisme producteur, ou par amplification PCR.

3.2.2.1. Par restriction de l'ADN génomique

La restriction de l'ADN génomique de l'organisme où le gène d'intérêt a été identifié est la méthode classiquement utilisée pour obtenir des fragments contenant ce gène en vue d'un TAR cloning. Afin de mettre en œuvre cette méthode, il faut préalablement identifier une ou des enzymes de restriction capables de digérer l'ADNg total sans altérer la séquence du gène d'intérêt. Dans le cas du gène *astS*, la taille du gène considéré (17 kb) génère un nombre très important de sites reconnus par des enzymes de restriction. Cet élément limite les possibilités de choix d'une enzyme capable de restreindre l'ADN génomique sans altérer l'intégrité de la séquence *astS*. En analysant le génome de *C. asteris*, au niveau du cluster contenant le gène *astS* (54 kb), seules deux enzymes peuvent être utilisées pour restreindre de part et d'autre du gène sans interrompre la séquence codante : *SgrDI* et *FseI* (**Figure 5**).



Figure 5 : Position des sites de restriction reconnus par les enzymes *FseI* (rose), *SgrDI* (vert) et *EcoRI* (blanc) au sein de la région génomique (54 kb) où se situe le gène *astS*.

Un faible nombre de sites de restriction reconnus par *FseI* et *SgrDI* ont été détectés au sein du génome de *C. asteris* (respectivement 521 et 1089 sites). L'enzyme *SgrDI*,

généralisant davantage de fragments, a donc été utilisée pour restreindre l'ADNg. Le profil de restriction obtenu après migration sur gel d'agarose est présenté **Figure 6A**.

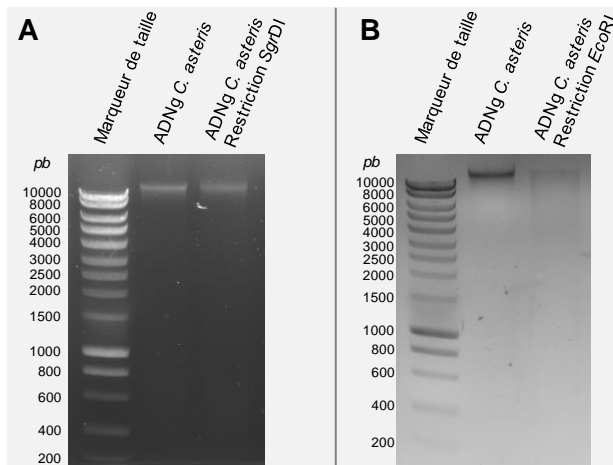


Figure 6 : Profil de migration obtenu avant et après restriction de l'ADNg de *C. asteris* en utilisant les enzymes de restriction *SgrDI* (A) et *EcoRI* (B).

Les profils obtenus pour l'ADNg de *C. asteris* avant et après restriction avec *SgrDI* sont similaires. L'utilisation de cette enzyme n'a permis d'obtenir qu'une restriction très partielle de l'ADNg et de générer principalement des fragments de tailles supérieures à 10 kb.

Les rendements de transformation par TAR cloning sont négativement impactés lorsque des fragments d'ADN de grande taille sont co-transformés avec le plasmide linéarisé. L'utilisation de l'ADNg restreint par *SgrDI* ne permet pas de réaliser un TAR cloning dans des conditions optimales. Une autre enzyme a donc été considérée pour la restriction de l'ADNg de *C. asteris*.

L'enzyme *EcoRI*, dont un seul site de restriction est présent au sein du gène *astS*, a été sélectionnée. Le génome de *C. asteris* présente davantage de site de restriction pour *EcoRI* (6741 sites) que pour *SgrDI* (1089 sites), générant ainsi des fragments plus nombreux et de taille plus réduite. En revanche, la présence d'un site de restriction *EcoRI* au sein du gène d'intérêt impose de réaliser uniquement une restriction partielle afin de limiter l'altération du gène *astS*. La restriction partielle de l'ADNg de *C. asteris* par *EcoRI* a été obtenue en diminuant le temps d'incubation avec l'enzyme de 30 à 5 min. Le profil de restriction obtenu après migration sur gel d'agarose est présenté **Figure 6B**. La présence d'un smear dans le profil de l'ADNg restreint par *EcoRI* témoigne de la présence d'une multitude de bandes de tailles diverses. L'utilisation d'*EcoRI* a permis d'obtenir une meilleure restriction de l'ADNg que celle obtenue via l'enzyme *SgrDI*. Il n'est néanmoins pas possible d'évaluer le niveau de conservation de la séquence *astS* à l'issue de cette restriction.

3.2.2.2. Par amplification PCR

En utilisant l'approche par amplification PCR, le gène *astS* amplifié peut être obtenu à des concentrations plus élevées et avec un haut taux de pureté. Au contraire de l'approche par restriction, le gène peut être purifié et n'est pas inclus avec l'ensemble des fragments de l'ADNg. Cependant, la grande taille du gène à amplifier (17 kb) nécessite le développement d'un protocole de PCR particulier. Plusieurs ADN polymérase ont donc été évaluées pour réaliser cette amplification, à partir de l'ADNg de *C. asteris*. Les produits PCR obtenus à l'aide des différentes polymérase ont été déposés sur gel d'agarose et comparés avec leur équivalent négatif où l'ADNg n'a pas été ajouté au mélange de la PCR (**Figure 7A**).

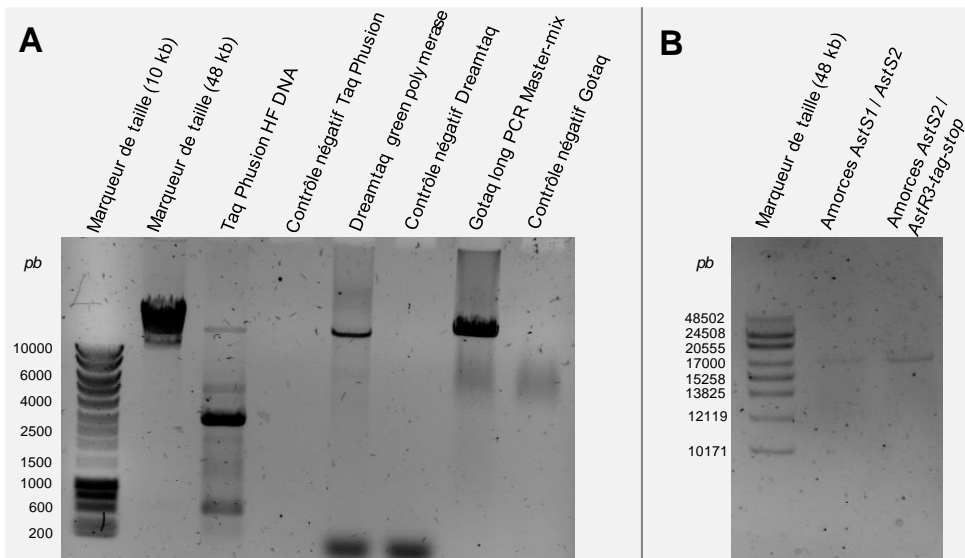


Figure 7 : Mise au point d'un protocole pour l'amplification et la purification du gène *astS* (17 kb). **A.** Résultat de l'amplification en utilisant différentes polymérase avec le couple d'amorces *AstS1/AstS2*. **B.** Résultat de l'amplification en utilisant la GoTaq long PCR Master-mix avec les couples d'amorces *AstS1/AstS2* et *AstS2/AstR3-tag-stop*, après mise au point d'un protocole de migration sur gel d'agarose adapté aux fragments de grande taille.

Un fragment de taille attendue a pu être obtenu à la suite de l'amplification PCR avec les polymérase Taq Phusion, DreamTaq, et GoTaq Long PCR. La Taq Phusion génère, en plus du fragment de 17 kb, de nombreux fragments amplifiés de manière non-spécifique. Les deux autres polymérase, la DreamTaq et la GoTaq Long PCR, ont permis d'amplifier de manière spécifique le fragment de 17 kb attendu. Les concentrations obtenues en utilisant la GoTaq Long PCR sont plus importantes que celles obtenues à partir de la DreamTaq (respectivement 39 et 9 ng/ μ L). Le protocole de migration (dépôt de 5 μ l sur gel d'agarose à 0.8 % Tris acétate EDTA, migration 30 min à 100 V) semble inadapté pour la séparation des fragments de grande taille, avec notamment une mauvaise séparation des 8 bandes contenues dans le mélange de marqueur de taille (48 kb).

La méthode de migration a donc été modifiée afin de favoriser la migration des fragments de grande taille. Le marqueur de taille a été dilué 5 fois dans H₂O (dépôt de 5 µL) et déposé sur un gel d'agarose à 0.3 % Tris acétate EDTA pour une migration de 2 h à 50 V. La séparation correcte des 8 bandes contenues dans le marqueur de taille (48 kb) a permis d'estimer correctement la taille du fragment amplifié par l'intermédiaire de la GoTaq long PCR en utilisant différentes amorces (**Figure 7B**). Le fragment amplifié présente bien une taille équivalente à 17 kb. Le fragment PCR généré à partir du couple d'amorces *AstS2* / *AstR3-tag-stop* (amorces permettant la dénaturation du codon stop de *astS* afin d'insérer une séquence FLAG dans le même cadre de lecture ouvert) a été purifié afin d'être co-transformé chez *S. cerevisiae* avec un plasmide linéarisé (TAR Cloning).

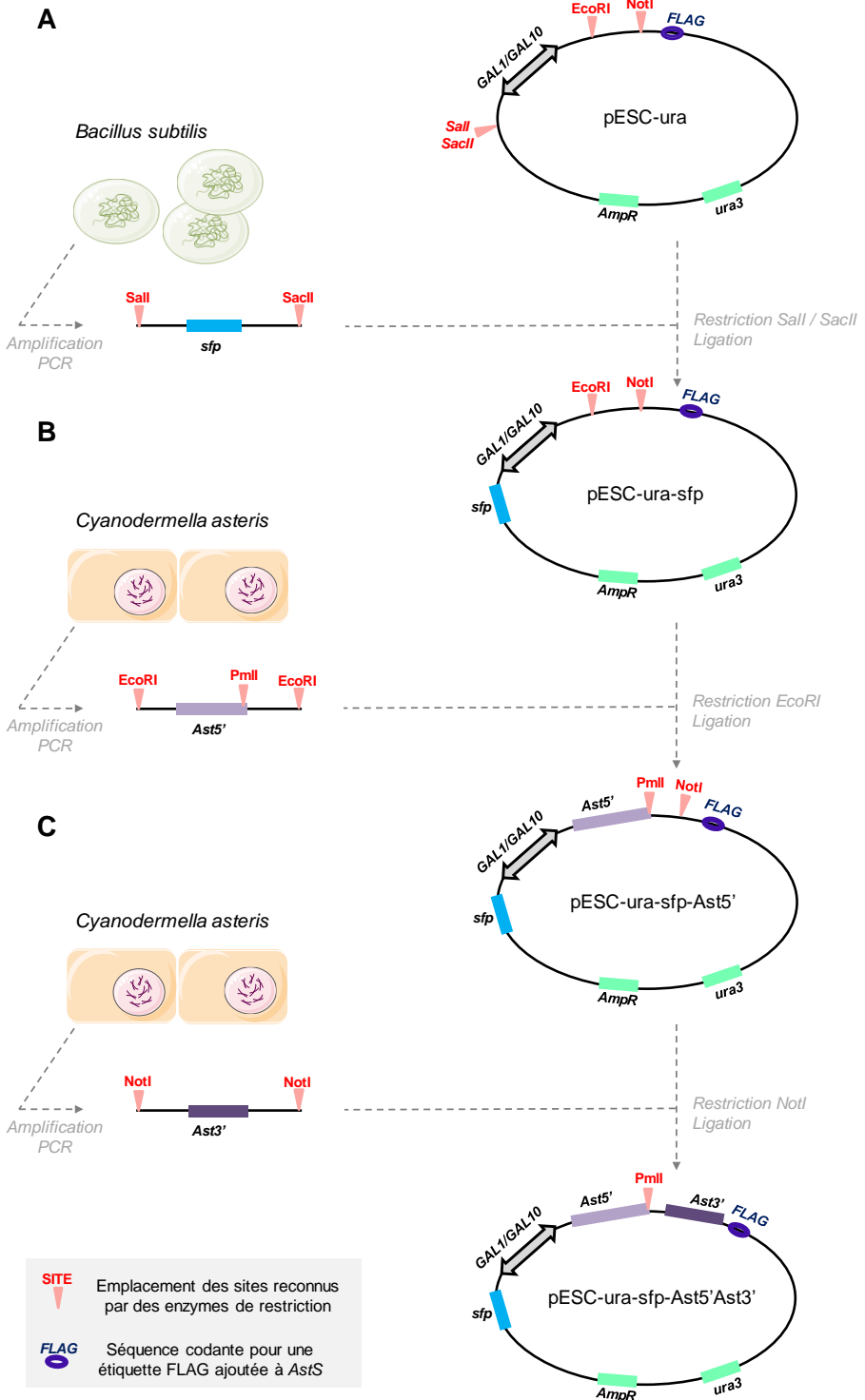
3.3. Expression hétérologue de *astS* (17 kb) chez *S. cerevisiae*

Cette section détaille les différentes approches qui ont été développées afin de mettre en place une expression hétérologue du gène *astS* chez *S. cerevisiae*. L'insertion de la séquence *astS* a été réalisée en utilisant la méthode du TAR cloning, soit en co-transformant avec la séquence *astS* obtenu par amplification PCR, soit avec des fragments d'ADN obtenus par restriction de l'ADNg de *C. asteris* (par *SgrDI* ou *EcoRI*). Dans un premier temps, la séquence *astS* a été clonée, sous contrôle d'un promoteur inductible au galactose, dans le plasmide pESC-ura. La séquence *astS* insérée dans ce plasmide a ensuite été libérée par restriction afin d'être transférée, sous contrôle d'un promoteur constitutif, dans un plasmide d'expression (pESC-ura ou pYX112).

3.3.1. Obtention des différents transformants de *S. cerevisiae*

Deux plasmides différents ont été utilisés pour réaliser l'expression hétérologue chez *S. cerevisiae*. Le premier est le plasmide pESC-ura qui est un vecteur multicopie équipé d'un promoteur GAL1 / GAL10 induit en présence de galactose. Le second, le plasmide pYX112, possède des séquences centromériques lui conférant des propriétés proches de celles des chromosomes levuriens, ainsi qu'un promoteur TPI (triose phosphate isomérase) qui est un promoteur constitutif fort de *S. cerevisiae*. Les deux plasmides sont équipés des gènes *AmpR* (induisant une résistance à l'ampicilline) et *URA3* (complémentant l'auxotrophie en uracile) permettent de sélectionner les transformants ayant intégrés le plasmide respectivement chez *E. coli* (en présence d'ampicilline) et *S. cerevisiae* Δ *ura3* (en absence d'uracile).

À partir de ces plasmides pESC-ura et pYX112, deux plasmides ont été assemblés : pESC-ura-sfp-*Ast5'*-*Ast3'* et pYX112-*Ast5'*-*Ast3'*. Les différentes étapes de construction mise en œuvre pour l'assemblage de ces deux plasmides sont présentées sur la **Figure 8** pour le plasmide pESC-ura-sfp-*Ast5'*-*Ast3'*, et **Figure 9** pour le plasmide pYX112-*Ast5'*-*Ast3'*.



◀ **Figure 8** : Différentes étapes nécessaires à l'assemblage du plasmide pESC-ura-sfp-Ast5'-Ast3', avec le clonage en premier lieu du gène *sfp* amplifié par PCR à partir de l'ADN de *B. subtilis* (A), puis de l'insertion du fragment *Ast5'* amplifié par PCR à partir de l'ADN de *C. asteris* (B), suivie par le clonage du fragment *Ast3'* également amplifié par PCR à partir de l'ADN de *C. asteris* (C).

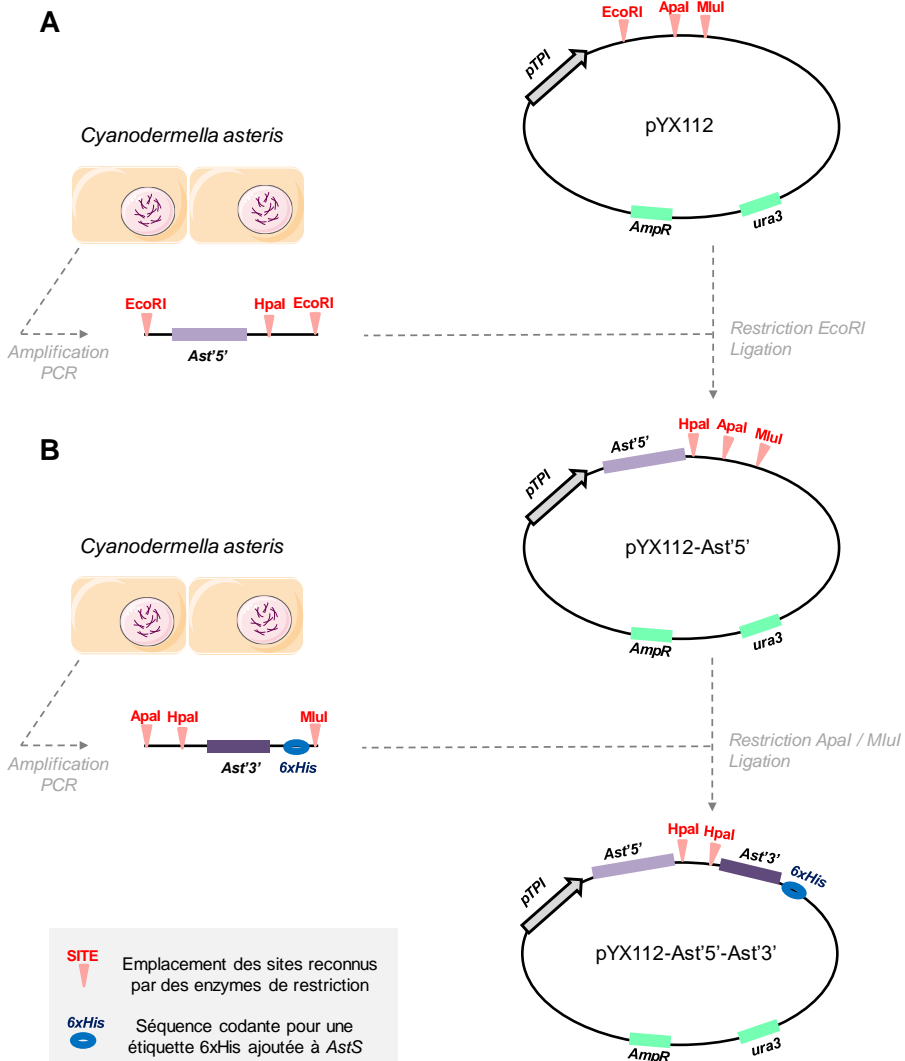


Figure 9 : Différentes étapes nécessaires à l'assemblage du plasmide pYX112-Ast5'-Ast3', avec l'insertion en premier lieu du fragment *Ast5'* amplifié par PCR à partir de l'ADN de *C. asteris* (A), suivi par le clonage du fragment *Ast3'* également amplifié par PCR à partir de l'ADN de *C. asteris* (B).

Le plasmide pESC-ura-sfp-Ast5'-Ast3' a été obtenu en clonant successivement les fragments *sfp* (815 pb amplifié à partir de l'ADN de *B. subtilis*) (**Figure 8A**), *Ast5'* (**Figure 8B**), et *Ast3'* (**Figure 8C**) (amplifiés à partir de l'ADN de *C. asteris*). Le fragment *Ast5'* est une séquence homologue sur 1373 pb à la partie 5' du gène *astS*. Le site de restriction *PmlI* présent à l'extrémité 3' de *Ast5'* (provenant de la séquence *astS*) sera utilisé par la suite pour la linéarisation du plasmide. Le fragment *Ast3'* est une séquence homologue sur 648 pb à la partie 3' du gène *astS*. Lors de l'amplification PCR, l'amorce *AstR3-tag-stop* utilisée a engendré une mutation sur le codon TGA (codon stop) le transformant en GGA (Gly). Cette mutation permet de placer le tag FLAG, initialement présent sur le plasmide pESC-ura, dans le même ORF que *Ast5'-Ast3'*.

En ce qui concerne le plasmide pYX112-Ast'5'-Ast'3', le fragment *Ast'5'* (1401 pb homologue à la partie 5' de *astS*) a été inséré dans un premier temps (**Figure 9A**), suivi par le fragment *Ast'3'* (779 pb homologue à la partie 3' de *astS*) (**Figure 9B**). Les deux fragments ont été amplifiés par PCR à partir de l'ADN de *C. asteris*. Un site de restriction *HpaI* a été inséré lors de l'amplification PCR à la partie 3' de *Ast'5'* par le biais de l'amorce rv *AstR5-restr-HpaI*. De même un site de restriction *HpaI* a été inséré à la partie 5' de *Ast'3'* par le biais de l'amorce fw *AstF3-HpaI*. Ces sites de restrictions *HpaI* seront utilisés par la suite pour linéariser le plasmide entre les fragments *Ast'5'* et *Ast'3'*. Lors de l'amplification PCR du fragment *Ast'3'*, un tag 6x HIS a été inséré au niveau de la partie 3' du fragment (dans le même ORF) par le biais de l'amorce *AstR3-tag-HIS-stop*.

Le plasmide pESC-ura-sfp-Ast5'-Ast3' linéarisé par *PmlI*, a été utilisé en premier lieu pour effectuer le TAR cloning. Afin de réaliser le TAR cloning, ce plasmide a été co-transformé dans *S. cerevisiae* DSY5, soit avec le fragment de 17 kb amplifié par PCR, soit avec de l'ADNg de *C. asteris* restreint par *SgrDI* ou *EcoRI*. Les différents transformants obtenus ont été analysés afin de vérifier l'insertion du gène *astS* (17 kb) au sein du plasmide par recombinaison homologue (**Tableau 7**).

Tableau 7 : Résultats obtenus après TAR cloning chez *S. cerevisiae*, en co-transformant avec une longue séquence *astS* isolée par des approches différentes.

Origine du fragment <i>AstS</i>	Nombre de clones testés	Nombre de transformant positif	Taux de transformation
Amplification PCR	10	9	90 %
Restriction ADNg de <i>C. asteris</i> par <i>SgrDI</i>	325	1	0.31 %
Restriction ADNg de <i>C. asteris</i> par <i>EcoRI</i>	92	3	3.26 %

En co-transformant avec le fragment de 17 kb obtenu par amplification PCR le taux de transformation est beaucoup plus élevé que les taux atteints en co-transformant avec de l'ADNg restreint. Le degré de pureté et la concentration du fragment constituent des éléments déterminants pour l'efficacité du TAR cloning. Le taux de transformation obtenu avec une restriction *EcoRI* est 10 fois supérieur à celui obtenu avec l'enzyme *SgrDI*. Le degré de restriction de l'ADNg (impactant le nombre et la taille des fragments) semble être un facteur déterminant en matière de TAR cloning. Le protocole de restriction partielle mis en place avec l'enzyme *EcoRI* a permis une meilleure restriction de l'ADNg de *C. asteris* tout en préservant suffisamment la séquence *astS* pour permettre son insertion dans le plasmide.

Un clone positif a été sélectionné parmi les transformants obtenus à partir du fragment généré par amplification PCR (pESC-sfp-AstS.PCR), et un parmi les transformants obtenus à partir de l'ADNg restreint par *EcoRI* (pESC-sfp-AstS.RESTR). La carte plasmidique de ces deux plasmides (théoriquement identiques) est présentée sur la **Figure 10**.

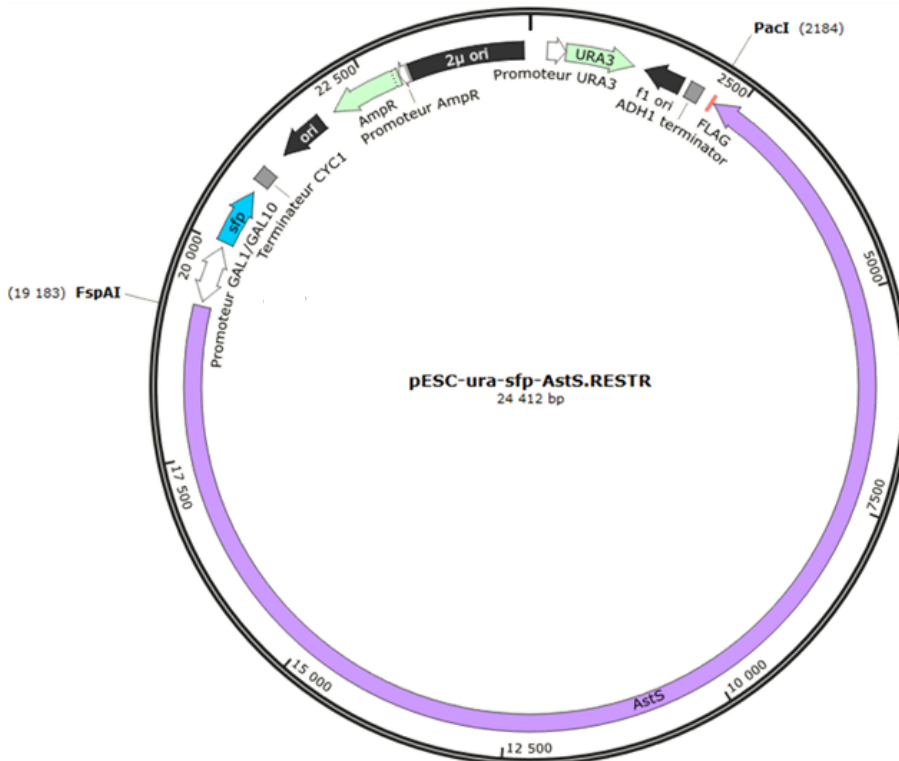


Figure 10 : Plasmide pESC-sfp-AstS.RESTR généré par TAR cloning dans *S. cerevisiae* par co-transformation du plasmide pESC-sfp-5'-3' et de l'ADNg de *C. asteris* restreint par *EcoRI*. Les positions des sites de restriction reconnus par les enzymes *FspAI* et *PacI* utilisées pour libérer par la suite le fragment afin de l'insérer dans le plasmide pYX112Ast'5'-Ast'3' sont indiquées.

Le plasmide pESC-sfp-AstS.RESTR a ensuite été restreint avec les enzymes *FspAI* et *PacI*, afin de libérer le fragment de 17 kb correspondant au gène *astS* et de pouvoir l'utiliser pour une co-transformation avec le plasmide pYX-Ast'5'-Ast'3' chez *S. cerevisiae*.

Après recombinaison homologue chez *S. cerevisiae* D5y5, le plasmide pYX112-AstS a été obtenu avec un taux de transformation de 100 % (sur 10 transformants évalués). La carte plasmidique de ce plasmide est présentée sur la **Figure 11**.

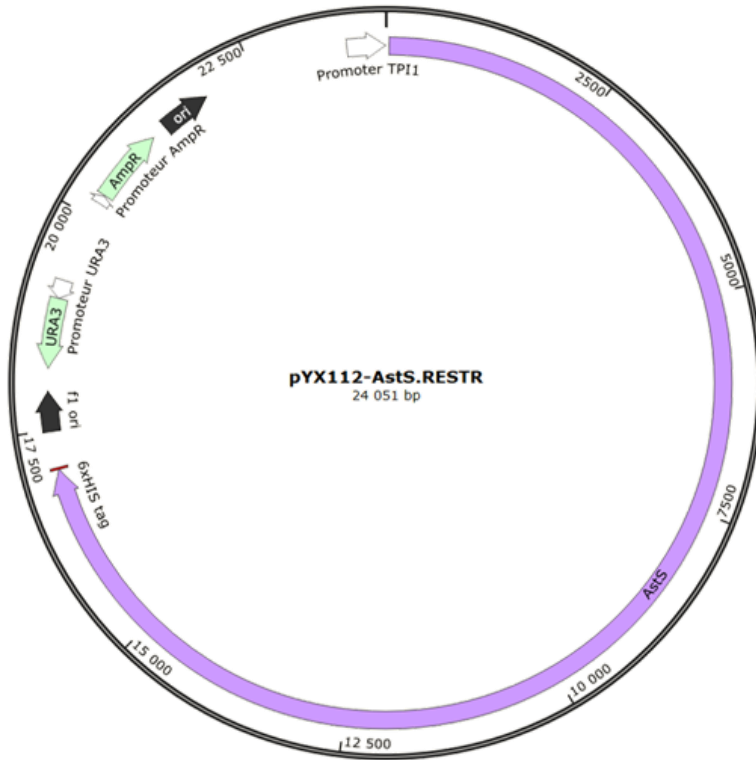


Figure 11 : Plasmide pYX112-AstS.RESTR généré par TAR cloning dans *S. cerevisiae* par co-transformation du plasmide pYX112-Ast'5'-Ast'3' et du fragment de 17 kb obtenu par restriction *FspAI* / *PacI* du plasmide pESC-sfp-AstS.RESTR.

Un transformant positif porteur du plasmide a été sélectionné pour la suite de l'étude. La **Figure 12** dresse le bilan des différentes constructions plasmidiques réalisées au cours de cette étude, et attribue ces plasmides aux souches de *S. cerevisiae* transformées correspondantes.

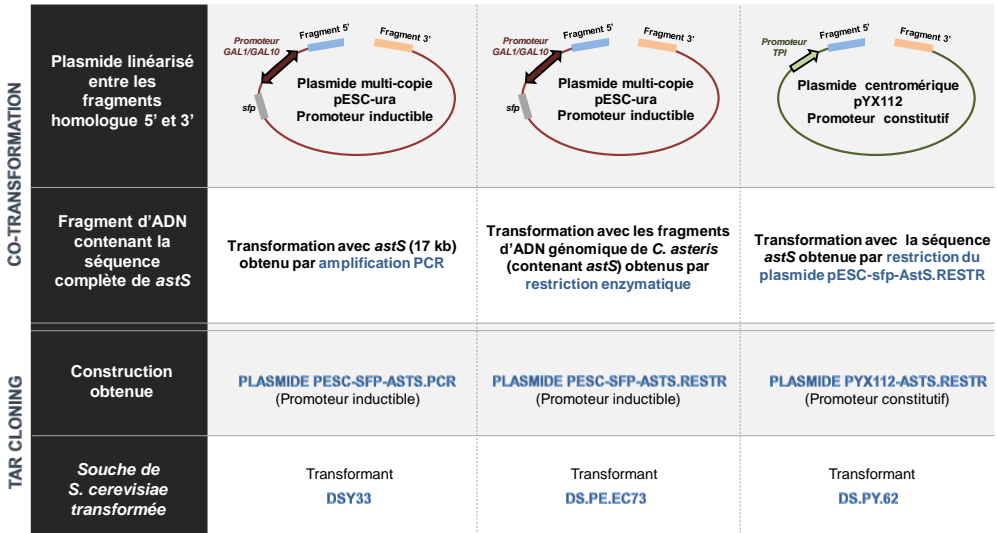


Figure 12 : Stratégie de clonage développée chez *S. cerevisiae*, avec l'insertion de *astS* (obtenu par restriction ou par amplification PCR de l'ADNg de *C. asteris*) au sein du plasmide pESC-ura (promoteur inducible GAL1 / GAL10), puis le transfert de cette séquence de pESC-ura à pYX112 (promoteur constitutif TPI – Triose Phosphate Isomérase).

3.3.2. Séquençage des clones positifs sélectionnés

Les plasmides (pESC-sfp-AstS.PCR, pESC-sfp-AstS.RESTR, et pYX112-AstS.RESTR) ont été extraits de *S. cerevisiae*, puis transformés chez *E. coli*. Le taux de réplication de l'ADN plasmidique étant plus important chez *E. coli* que chez *S. cerevisiae*, une quantité suffisante d'ADN plasmidique (> 100 ng/μL) a pu être extraite et purifiée. Les plasmides purifiés ont ensuite été séquencés afin de contrôler l'absence de mutations dans la séquence *astS* insérée au sein des différents plasmides.

Aucune mutation n'a été détectée chez les plasmides pESC-sfp-AstS.RESTR et pYX112-AstS.RESTR obtenus à partir de fragments d'ADNg de *C. asteris* restreint par *EcoRI*.

En revanche, le plasmide pESC-sfp-AstS.PCR obtenu à partir d'un fragment amplifié par PCR présente 13 mutations réparties sur l'ensemble de la séquence *astS* (**Tableau 8**).

Tableau 8 : Résultats du séquençage du gène *astS* inséré dans le plasmide pESC-sfp-*AstS*.PCR. 13 mutations ont été détectées au sein de la séquence génétique dont deux mutations silencieuses (vert) et deux susceptibles d'impacter fortement la structure tridimensionnelle par inclusion de l'acide aminé cyclique proline (rouge).

N°	Position dans <i>astS</i> (pb)	Codons (ADN)		Acide aminé (AA)		Position de l'AA dans la NRPS	Module / Domaine concerné
		Intègre	Muté	Intègre	Muté		
1	1186	GTC	GCC	Val	Ala	396	Domaine A (Mod. 1)
2	2407	AAG	AGG	Lys	Arg	803	Inter-domaine A-T (Mod.1)
3	2571	TCT	CCT	Ser	Pro	858	Domaine T (Mod. 1)
4	2947	GAG	GGG	Glu	Gly	983	Domaine C (Mod. 2)
5	3642	TAT	CAT	Tyr	His	1215	Inter-domaine C-A (Mod. 2)
6	4874	GCA	GCG	Ala	Ala	1625	Domaine A (Mod. 2)
7	5863	CGC	CAC	Arg	His	1955	Domaine T (Mod. 2)
8	8693	GTC	GCC	Val	Ala	2865	Inter-domaine A-T (Mod. 3)
9	9507	TCT	CCT	Ser	Pro	3170	Domaine C (Mod. 4)
10	10044	GTT	ATT	Val	Ile	3349	Inter-domaine C-A (Mod. 4)
11	10093	AGC	AAC	Ser	Asn	3365	Inter-domaine C-A (Mod. 4)
12	11699	CGA	CGG	Arg	Arg	3900	Domaine A (Mod. 4)
13	15105	ATG	GTG	Met	Val	5036	Inter-domaine A-T (Mod. 5)

Les mutations 6 et 12 sont silencieuses et ne modifient pas l'acide aminé incorporé dans la séquence protéique. Elles n'ont donc aucun impact sur la structure de la synthétase. Les 11 autres mutations entraînent une modification de l'acide aminé incorporé au sein de la synthèse et donc potentiellement une modification de la conformation tridimensionnelle de la synthétase. En particulier, les mutations 3 et 9 entraînent le remplacement de la sérine, un acide aminé hydroxylé, par une proline, un acide aminé cyclique. La structure cyclique de la proline est un facteur connu induisant la rupture des structures de type feuillets β en empêchant l'établissement du réseau de liaison hydrogène nécessaire à leur maintien [36,37]. Les mutations

impliquant l'insertion d'une proline sont donc susceptibles d'impacter profondément la structure de la NRPS. L'impact de la mutation N° 9 positionnée dans le domaine de condensation du module 4 a été évaluée en se basant sur le modèle tridimensionnel généré précédemment (**Chapitre 3 – Section 2.4.1**). La prédiction des changements structuraux introduits par la substitution de l'acide aminé Ser par Pro a été réalisée par l'intermédiaire de l'outil Missense3D [27] (**Figure 13**).

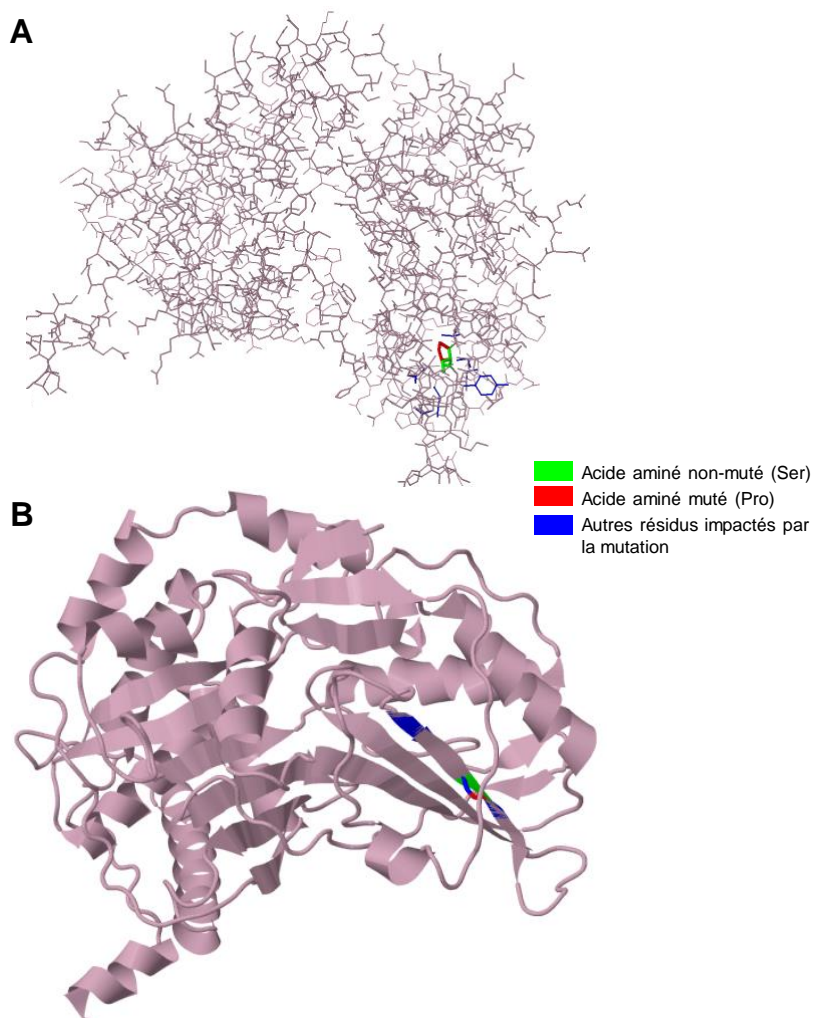


Figure 13 : Visualisation de la position de la mutation N° 9 (Ser3170Pro) au sein du domaine de condensation du module 4 et estimation de l'impact de cette mutation sur la structure du domaine via le logiciel Misense3D. **A.** Vue à l'échelle des liaisons atomiques. **B.** Vue de la structure tridimensionnelle et de l'organisation en hélices α / feuillets β .

La mutation Ser3170Pro a un impact local sur la conformation adopté par 5 autres résidus (en bleu - **Figure 13A**) et entraîne la disparition du feuillet bêta présent en absence de mutation (**Figure 13B**). Une telle modification structurale du domaine est susceptible d'entraîner une perte de l'activité de condensation et de compromettre l'assemblage du NRP par la synthétase. Étant donné le nombre important de mutations observées, et leur impact supposé sur l'intégrité de la structure, les travaux n'ont pas été poursuivis avec le transformant porteur du plasmide pESC-ura-sfp-AstS.PCR. Seuls les transformants ayant intégrés les plasmide pESC-ura-sfp-AstS.RESTR et pYX112-AstS.RESTR, où la séquence génétique est parfaitement conservée, ont été considérés pour la suite des analyses.

3.3.3. Évaluation de la transcription

Les souches *S. cerevisiae* DS.Pe.Ec73 et DS.Py.62 ont été cultivées dans un milieu minimum MAM supplémenté en Leu, His, Trp, et dépourvu de Ura pour maintenir la pression de sélection sur les transformants (*ura3* présent dans les plasmides pESC et pYX112). Le milieu a été supplémenté en galactose pour la culture de DS.Pe.Ec73 afin de permettre l'induction du promoteur GAL1 / GAL10. La souche DSY5 (contrôle négatif) a été cultivée dans des conditions identiques à celles des deux souches transformées, mais avec un milieu supplémenté en uracile.

Pour chaque culture, le niveau d'expression du gène *astS* a été évalué au début de la phase exponentielle de croissance (16 h) et à la fin de la phase exponentielle de croissance (36 h). Les cycles seuils (cycle à partir duquel la valeur de fluorescence mesurée est supérieure à celle du bruit de fond) obtenus pour chacun des échantillons analysés par qPCR sont présentés sur le **Tableau 9**.

Tableau 9 : Résultats obtenus en termes de Ct (cycle seuil) lors des qPCR réalisées, afin d'évaluer la présence du transcrite du gène *astS*, avec le couple d'amorces *qpAstS1 / qpAstS2* sur les ADNc synthétisés à partir des ARNm extraits des souches DSy5 (contrôle négatif), DS.Pe.Ec73, et DS.Py.62.

Souche	Milieu	Temps (h)	Ct
DSy5	MAM + Gal	16	37,1358
DSy5	MAM	16	35,3056
DSy5	MAM + Gal	36	32,3112
DSy5	MAM	36	33,7245
DS.Pe.Ec73	MAM + Gal	16	23,8896
			23,5908
DS.Pe.Ec73	MAM + Gal	36	21,9130
			25,8677
DS.Py.62	MAM	16	22,4790
			25,9919
DS.Py.62	MAM	36	23,9316
			27,1685

On distingue deux groupes de valeurs de cycle seuil. Les valeurs de Ct obtenues pour l'ensemble des échantillons provenant de la souche non-transformée DSy5 sont situées entre 32 et 38. A l'inverse, les échantillons provenant des souches transformées DS.Pe.Ec73 et DS.Py.62, présentent des Ct qui sont systématiquement inférieurs à 28.

Dans le cas d'une production hétérologue, la RT-qPCR est utilisée pour différencier des échantillons où le transcrite est potentiellement présent (souche transformée) vis-à-vis d'échantillon où le gène et son ARNm correspondant sont absents (souche non-transformée). La valeur 30 de Ct peut être considérée comme une valeur de coupure au-delà de laquelle les amplifications observées sont non-spécifiques [28].

Par conséquent, aucune amplification spécifique n'a été détectée à partir des ADNc extraits des cellules de DSy5. En revanche, la présence de l'ADNc correspondant à *astS* a été détectée chez DS.Pe.Ec73 et DS.Py.62 après 16 et 36 h de culture.

Afin de confirmer l'absence de contamination des échantillons avec de l'ADNg, une amplification PCR a été réalisée en utilisant les mêmes amorces sur les ARNs extraits de DS.Pe.Ec73 et DS.Py.62 (avant transcription inverse). Aucune amplification n'a été obtenue, confirmant la présence de l'ARN uniquement (données non-présentées). La transcription du gène *astS* est donc bien effective à la fois chez DS.Pe.Ec73 et DS.Py.62.

Afin de comparer les niveaux d'expressions entre ces deux souches (où *astS* est sous contrôle de promoteurs différents), les valeurs moyennes de $2^{-\Delta\Delta Ct}$ ont été calculées pour chacune des conditions (16 et 36 h de culture). Les valeurs relatives obtenues pour les échantillons provenant de la souche DS.Py.62, en fixant l'échantillon provenant de la souche DS.Pe.Ec73 comme condition contrôle (à 1), sont présentées sur la **Figure 14**.

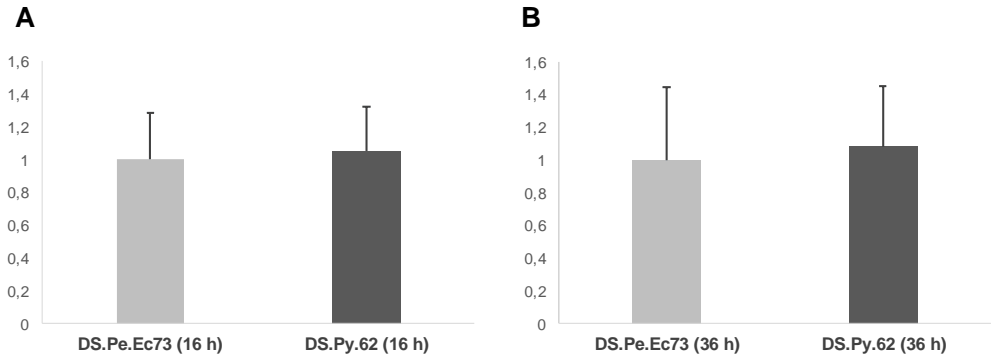


Figure 14 : Valeurs moyennes relatives de $2^{-\Delta\Delta Ct}$ obtenues pour les échantillons provenant de la souche DS.Py.62 en fixant les échantillons provenant de la souche DS.PE.Ec73 en tant que condition contrôle, à 16 h de culture (**A**) et à 36 h de culture (**B**). Les moyennes et écarts types ont été calculés sur des cultures réalisées en duplicata.

Aucune variation significative n'a pu être mise en évidence concernant les niveaux d'expression du transcrite de *astS* entre la souche DS.Pe.Ec73 (promoteur GAL1 / GAL10 inductible) et la souche DS.Py.62 (promoteur TPI constitutif).

3.3.4. Analyses protéomiques

Les protéines totales ont été extraites à partir des souches DSy5 (contrôle négatif), DS.Pe.Ec73, et DS.Py.62 afin de déterminer si le transcrite du gène *astS* est ensuite traduit en protéine chez les transformants. Les protéines extraites des différentes souches (après 24 h de culture) ont ensuite été analysées par SDS-Page (**Figure 15**).

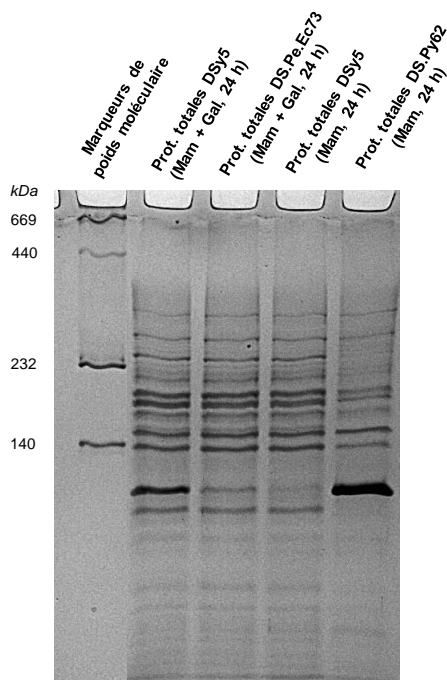


Figure 15 : Profil de migration des protéines totales extraites à partir des cellules de DSy5, DS.Pe.Ec73, et DS.Py62, après migration dans un gel SDS-Page. Les marqueurs de poids moléculaire sont constitués de thyroglobuline (669 kDa), de ferritine (440 kDa), de catalase (232 kDa), et de lactate déshydrogénase (140 kDa).

Aucune bande spécifique correspondant à l'astine synthétase (520 kDa) n'a pu être détectée entre 440 et 669 kDa après migration. Afin de résoudre d'éventuels problèmes de solubilisation de la NRPS lors de l'extraction des protéines, la même expérience a été réalisée en utilisant un tampon d'extraction modifié (Tampon B). Plusieurs études rapportant une localisation péri-membranaire des NRPSs [38,39], ce tampon alternatif a été formulé afin de faciliter la solubilisation de protéines hydrophobes, ainsi que des protéines associées aux structures membranaires. La migration sur gel des protéines extraites à partir de ce tampon n'a pas non plus permis de mettre en évidence une bande spécifique à 520 kDa correspondante à la NRPS AstS ni chez DS.Pe.Ec73, ni chez DS.Py.62 (données non présentées).

Les étiquettes ajoutées à la partie C-terminale de la NRPS (FLAG pour la synthétase exprimée par DS.Pe.Ec73 et 6x HIS pour celle exprimée par DS.Py.62) ont été utilisées pour la détection de la NRPS par western blot. La méthode d'immunoblotting permet la détection de protéines plus faiblement concentrées (jusqu'à l'échelle du pico-gramme) et parfois non visible sur gel SDS Page. Un anticorps anti-DYKDDDDK (FLAG) et un anticorps anti-poly-HIS ont été utilisés pour incuber les membranes de nitrocellulose comportant des protéines issues respectivement de DS.Pe.Ec73 et de DS.Py.62. Après révélation, aucune bande correspondant à la NRPS AstS n'a pu être détectée pour les deux western blot réalisés (données non présentées).

L'étiquette poly-histidine présente sur la partie C-terminale de la NRPS exprimée par DS.Py.62 a également été utilisée pour purifier la synthétase sur colonne de purification HisTrap. Après élution de la fraction protéique, dialyse, et concentration (x 100), aucune bande correspondante à la NRPS AstS n'a pu être détectée (données non présentées).

La traduction de l'ARNm de l'*astS* n'a pas pu être mise en évidence chez les deux transformants considérés. Il est possible que la séquence de la NRPS soit trop large pour être traduite ou que la quantité de protéine produite soit en dessous du seuil de détection. L'hypothèse d'une instabilité du complexe NRPS chez *S. cerevisiae*, voire d'une dégradation par l'action de protéases, n'est pas à exclure.

3.4. Expression hétérologue de AstS tronquée (2 modules et C_{term}) chez *Y. lipolytica*

L'approche du TAR cloning développée chez *S. cerevisiae* pour le clonage du gène *astS* (17 kb) est difficilement transposable chez *Y. lipolytica* où l'efficacité de la recombinaison homologue est moindre. Bien que d'autres outils soient disponibles pour le clonage de grands fragments d'ADN [40–42], le gène *astS* n'a pas été cloné dans son intégralité au sein de *Y. lipolytica*. Cette levure n'ayant jamais été utilisée précédemment pour réaliser une production hétérologue de NRP, seule une partie du gène codant pour la NRPS AstS a été clonée. Cette stratégie permettra d'évaluer la capacité de *Y. lipolytica* à exploiter une voie de synthèse de type non-ribosomique à partir d'une NRPS de deux modules. La séquence clonée au sein du génome de *Y. lipolytica* code pour les deux premiers modules de la synthétase suivis, dans le même cadre de lecture, par le domaine de condensation terminal (**Figure 1**).

3.4.1. Préparation des plasmides et obtention des transformants

Le fragment du gène *astS* codant pour les 2 premiers modules (du codon d'initiation jusqu'au début du deuxième domaine de condensation délimité par l'analyse antiSMASH- **Chapitre 3 – section 2.2**) a été amplifié à partir du plasmide pYX112.AstS.RESTR à l'aide des amorces *BamHI-Ast6kb-fw* et *Ast6kb-sfiIA-sfiIB-AvrII-rv*. La première amorce ajoute un site de restriction *BamHI* sur la partie 5' du fragment, tandis que la seconde ajoute 3 sites de restrictions successifs *SfiA*, *SfiB*, et *AvrII*. Le fragment de 6012 pb amplifié a été restreint avec les enzymes *BamHI* / *AvrII*, et inséré dans le plasmide RIP191 linéarisé avec des enzymes identiques (**Figure 16A-B**).

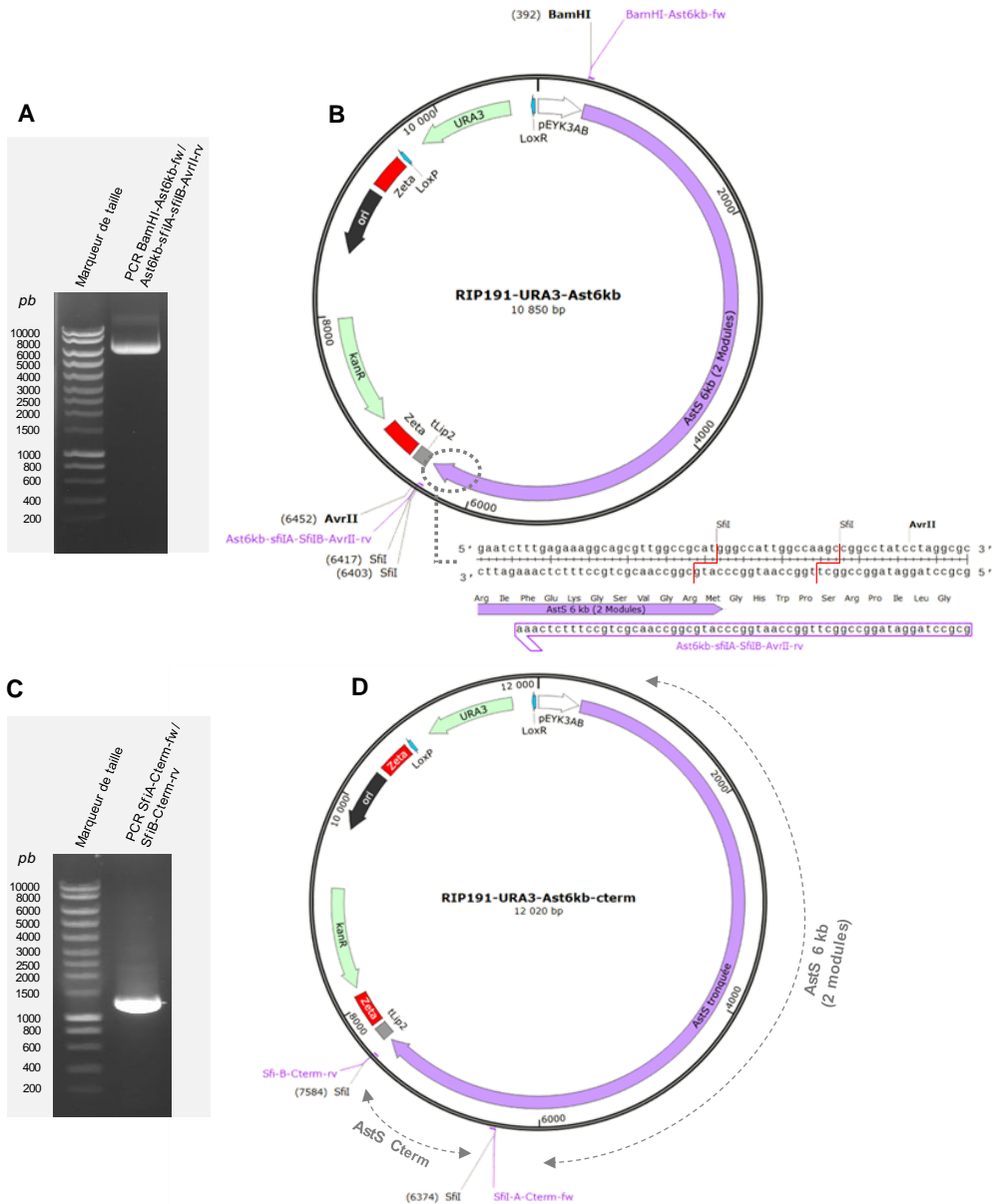


Figure 16 : Assemblage des séquences codantes pour les 2 premiers modules de l'AstS, suivis par le domaine C_{term}, au sein du plasmide RIP191-URA3. **A.** Gel d'électrophorèse après migration du fragment de 6012 pb amplifié par PCR avec le couple d'amorces *BamHI-Ast6kb-fw / Ast6kb-sfiA-SfiB-AvrII*. **B.** Carte plasmidique de RIP191-URA3-Ast6kb après insertion du fragment de 6 kb amplifié par PCR. Les sites de restrictions *SfiI* requis pour l'insertion du domaine C_{term} sont localisés au niveau de l'extrémité 3' du fragment Ast6kb. **C.** Gel d'électrophorèse après migration du fragment de 1229 pb amplifié par PCR avec le couple d'amorces *SfiA-Cterm-fw / SfiB-Cterm-rv*. **D.** Carte plasmidique de RIP191-URA3-Ast6kb-Cterm après insertion du fragment de 1,2 kb amplifié par PCR.

La séquence codante pour la partie C_{term} de l'AstS (1229 pb) (délimitée par l'analyse anti-SMASH – **Chapitre 3 – Section 2.2**) a été amplifiée par PCR à partir du plasmide pYX112.AstS.RESTR à l'aide des amorces *SfiA-Cterm-fw* et *SfiB-Cterm-rv* (**Figure 16D**). Le tag 6xHIS présent sur la partie 3' du domaine C_{term} inséré dans pYX112.AstS.RESTR a été conservé lors de l'amplification PCR. *SfiI* est une enzyme capable de reconnaître les sites GGCCNNNN|NGGCC (où « | » constitue le site de coupure, et « N » n'importe quel nucléotide) [43]. Des sites de restriction *SfiI* comportant des séries de nucléotides « N » différentes ont été insérés par le biais des amorces sur les parties 5' et 3' du C_{term} amplifié. En insérant des bases nucléotidiques distinctes au niveau des deux sites *SfiI*, deux extrémités cohésives 3' différentes ont été générées après restriction via *SfiI* (5'-ATG-3' pour l'extrémité 5', et 5'-AGC-3' pour l'extrémité 3') (**Figure 16B**). Ces extrémités cohésives, homologues à celles générées par restriction *SfiI* du plasmide RIP191-URA3-Ast6kb, ont permis d'insérer le fragment C_{term} dans le même sens 5'-3', et dans le même cadre de lecture, que le fragment Ast6kb (**Figure 16 C-D**).

Le plasmide RIP191-URA3-Ast6kb-Cterm obtenu a ensuite été séquencé, afin de confirmer l'absence de mutation et l'insertion correcte du fragment Ast6kb et C_{term} dans le même ORF. La cassette Zeta – LoxP – URA3 – LoxR – pEYK300A3B – AstS6kb – Cterm – Termineur Lip2 – Zeta a ensuite été intégrée dans le génome de *Y. lipolytica* par recombinaison homologue avec les séquences zeta. Un transformant positif ayant intégré la séquence Ast6kb-C_{term}, nommé *astS.Tr* (pour AstS tronquée), a été sélectionné (clone B3).

Une seconde construction plasmidique a été assemblée en parallèle. Le gène *npgA*, codant pour une phosphopantéthéine transférase, a été amplifié par PCR en utilisant le couple d'amorces *HindIII-NpgA-fw* / *AvrII-NpgA-rv* à partir de l'ADN génomique d'*Aspergillus nidulans*. Le fragment de 1059 pb a ensuite été inséré au sein du plasmide RIP222-LEU2 générant ainsi le plasmide RIP222-LEU2-NpgA (**Figure 17**).

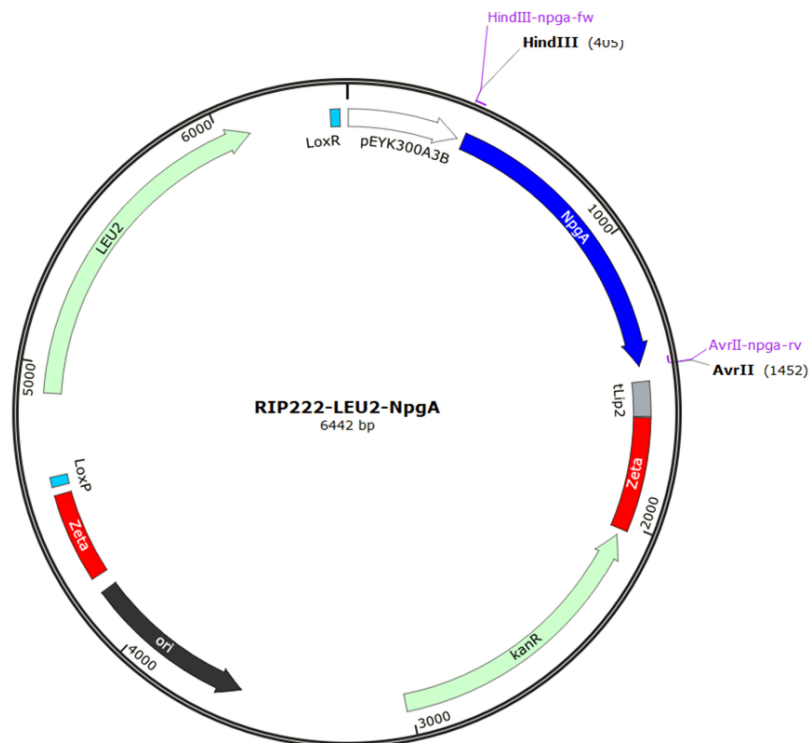


Figure 17 : Obtention du plasmide RIP222-LEU2-NpgA par ligation entre RIP222-LEU2 et le fragment PCR NpgA de 1059 pb (restreint par l'action des enzymes *HindIII* et *AvrII* dont l'emplacement des sites de restriction est indiqué).

Aucune mutation n'a été identifiée sur le séquençage du plasmide RIP222-LEU2-NpgA obtenu. La cassette Zeta – LoxP – LEU2 – LoxR – pEYK300A3B – NpgA – Terminateur Lip2 – Zeta a ensuite été intégrée dans le génome du clone *Y. lipolytica* B3. Un transformant positif ayant intégré dans son génome les séquences *npgA* et *astS.Tr* sous contrôle du promoteur pEYK300A3B, a été sélectionné (clone B3R2).

3.4.2. Évaluation de la transcription

Les séquences *npgA* et *astS.Tr* ont été intégrées dans le génome de *Y. lipolytica* sous contrôle du promoteur pEYK300A3B. Ce promoteur est induit par la présence d'erythritol dans le milieu de culture. La souche RIY261 de *Y. lipolytica* transformée étant mutée pour le gène *EYD1*, l'erythritol prélevé par la cellule n'est pas consommé, permettant de maintenir une induction continue.

La souche *Y. lipolytica* non-transformée (RIY261) et la souche transformée B3R2 (*astS.Tr+*, *npgA+*) ont été cultivées sur milieu YPD supplémenté en erythritol. L'utilisation du milieu YPD (milieu riche) permet d'obtenir des taux de croissance plus élevés. Afin de déterminer les niveaux de transcription des gènes *astS.Tr* et *npgA*, une RT-qPCR a été réalisée sur les ADNc obtenus à partir de la souche B3R2, à 8, 12, 18 et 24 h de culture. Les ADNc obtenus à partir des échantillons de la souche RIY261

ont été considérés en tant que témoin négatif. Les résultats obtenus en termes de Ct pour chacun des échantillons analysés sont présentés sur le **Tableau 10**.

Tableau 10 : Résultats obtenus en termes de Ct (cycle seuil) lors des RT-qPCR réalisées à partir des ADNc extraits des souches B3R2 et RIY261 en utilisant le couple d'amorces *qpAend1* / *qpAend2* (**A**) et *qp-npg1-fw* / *qp-npg1-rv* (**B**).

A <i>qpAend1</i> / <i>qpAend2</i>				B <i>qp-npg1-fw</i> / <i>qp-npg1-rv</i>			
Souche	Milieu	Temps (h)	Ct	Souche	Milieu	Temps (h)	Ct
RIY261	YPD	8	31,0963	RIY261	YPD	12	35,9398
RIY261	YPD	12	31,7494	B3R2	YPD	8	25,5927
RIY261	YPD	18	31,6085	B3R2	YPD	12	24,4338
RIY261	YPD	24	31,0288	B3R2	YPD	18	24,6977
B3R2	YPD	8	20,7923	B3R2	YPD	24	24,5117
B3R2	YPD	12	19,6474				
B3R2	YPD	18	20,4858				
B3R2	YPD	24	19,4277				

Les valeurs de Ct obtenues à partir des échantillons provenant de la souche RIY261, pour les deux couples d'amorces, sont supérieures à 31. Chez les échantillons provenant de la souche B3R2, les Ct sont inférieurs à 21 pour les amorces relatives à *astS*, et inférieurs à 26 pour celles relatives à *npgA*.

Aucune amplification spécifique n'a été obtenue ($Ct > 30$) pour les échantillons obtenus à partir de la souche RIY261 (non transformée). En revanche, les Ct compris entre 19 et 26, observés pour les échantillons obtenus à partir de la souche B3R2 (*astS.Tr+*, *npgA+*), sont caractéristiques d'une amplification spécifique. La présence du transcrite *astS* et *npgA* est donc confirmée chez la souche transformée B3R2 tout au long de la phase exponentielle de croissance (de 8 à 24 h de culture).

3.4.3. Analyses protéomiques

Afin de mettre en évidence la traduction du transcrite *astS.Tr*, les protéines ont été extraites à partir de la souche B3R2 après 8 et 24h de culture. Le profil protéique de cette souche a été comparé avec celui de la souche non transformée (RIY261) après migration sur un gel SDS-Page (**Figure 18**).

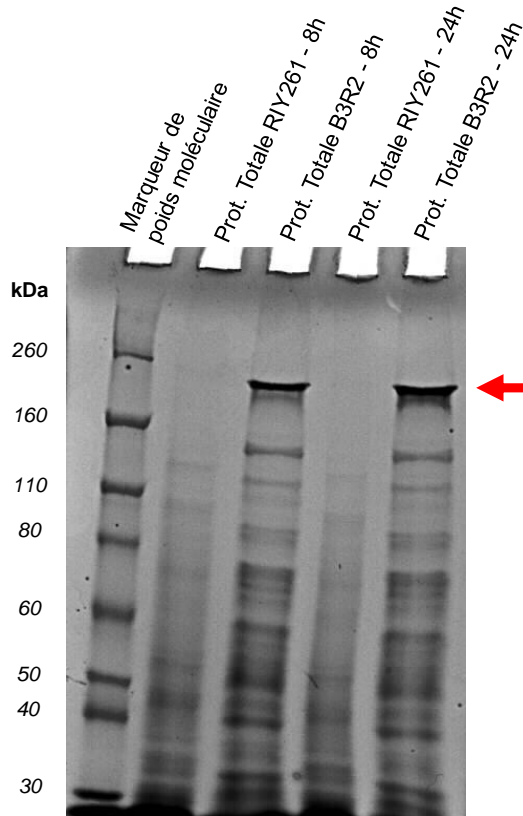


Figure 18 : Profil de migration sur SDS-Page des lysats protéiques extraits à partir des cultures de R1Y261 (contrôle négatif) et B3R2 (souche transformée *astS.Tr+*, *npgA+*) après 8 et 24 h de culture sur milieu YPD supplémenté en erythritol.

Entre 160 et 260 kDa, une bande est présente uniquement au sein des profils protéiques correspondants à la souche B3R2, et pas sur ceux correspondants à R1Y261. La taille de cette bande a été estimée à 210 kDa, soit légèrement inférieure à celle attendue pour la protéine correspondante à la protéine AstS.Tr (262 kDa).

La bande a été prélevée pour réaliser des analyses complémentaires visant à confirmer la présence de la protéine AstS.Tr. Un traitement de réduction-alkylation-digestion (trypsine) a été appliqué afin de digérer la protéine intacte (entière) pour obtenir des fragments peptidiques pouvant être analysés en spectrométrie de masse. L'analyse a permis de mettre en évidence 5 fragments peptidiques appartenant à la séquence protéique de la protéine AstS.Tr (**Tableau 11**).

Tableau 11 : Fragments peptidiques appartenant à la séquence protéique de AstS.Tr identifiés par spectrométrie de masse. (M* : oxydation +15,99 sur le résidu M).

N°	Peptide	Masse (Da)	m/z	Position	
				Début	Fin
1	K.TPEDSGK.I	732,3290	733,3817	1660	1666
2	R.GAAEH.M	483,2077	484,9651	1862	1866
3	R.KIESQM*APEPKVDILRR.V	2025,1040	675,99999	803	819
4	R.LIPEEIK.W	840,4956	841,5922	631	637
5	L.LYHGRAATQIK.I	1256,6990	629,3397	1750	1760

Les 5 fragments peptidiques identifiés sont localisés au sein de la séquence protéique appartenant aux 2 premiers modules (Position 1 à 1999). Aucun fragment peptidique appartenant à la partie C_{term} n'a été détecté.

Afin de valider les résultats obtenus lors de l'analyse par spectrométrie de masse, et de vérifier la présence du domaine C_{term} au sein de cette protéine AstS.Tr, un western blot a été réalisé (**Figure 19**).

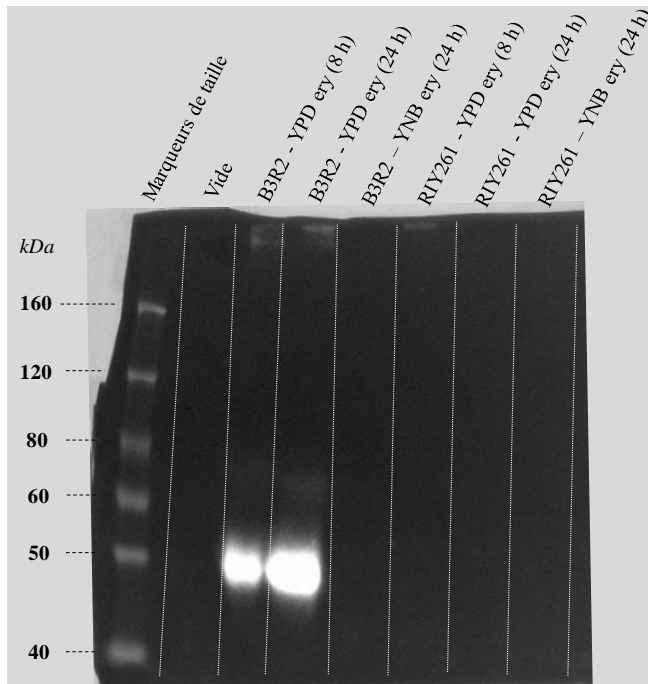


Figure 19 : Révélation du western blot réalisé avec des anticorps dirigés contre le tag 6x HIS sur des protéines extraites à partir des souches B3R2 et RIY261 cultivées sur YPD ou YNB pendant 8 ou 24 h.

Les anticorps utilisés pour le western blot étaient dirigés spécifiquement contre le tag 6xHis positionné au niveau de la partie C-terminale du domaine de condensation terminal. Un signal de forte intensité a été détecté entre 40 et 50 kDa pour le transformant B3R2 cultivé sur milieu YPD. Aucun signal n'a été mis en évidence pour la souche non-transformée RIY261 ou pour B3R2 cultivé sur milieu YNB.

Le domaine C_{term}, isolé du reste de la protéine AstS.Tr, possède un poids moléculaire de 45 kDa. Le signal observé sur le western blot correspond à une protéine porteuse d'un His-tag d'une taille similaire à celle du domaine C_{term}.

Aucun fragment peptidique appartenant au domaine C_{term} n'ayant été détecté lors de l'analyse par spectrométrie de masse, il semble que la protéine AstS.Tr soit traduite, puis scindée en deux parties au niveau de la zone inter-domaine entre la fin du module 2 et le début du domaine C_{term}.

3.4.4. Évaluation de la production du NRP : L-2Abu-LPro

Différentes analyses ont ensuite été conduites afin de déterminer si la synthétase Ast.Tr conserve son activité malgré la potentielle scission entre les deux premiers modules et le domaine C_{term}. Dans l'hypothèse où Ast.Tr est toujours fonctionnelle, un dipeptide L-2Abu – L-Pro sera assemblé au sein de la souche B3R2. Une méthode d'analyse par LC-MS a été développée pour le dosage de ce dipeptide (**Figure 20**).

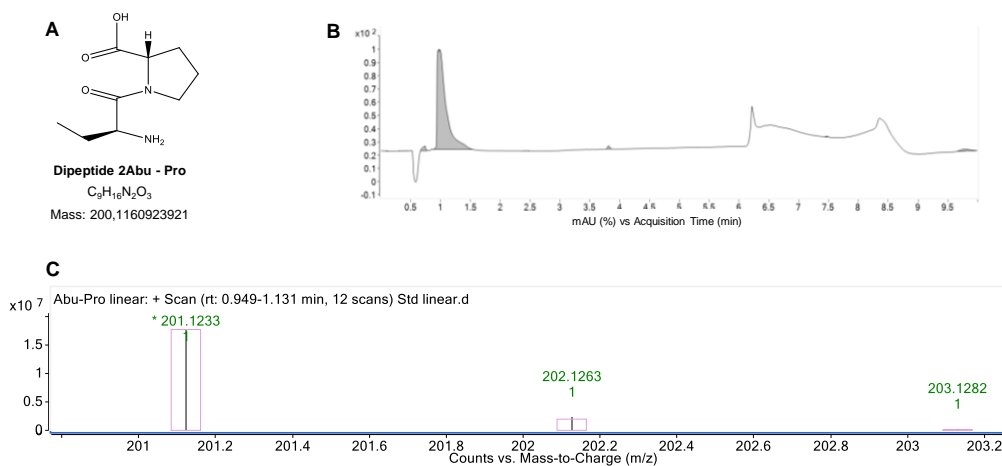


Figure 20 : Mise au point d'une méthode de dosage pour le dipeptide L-2Abu – L-Pro (**A**) par l'intermédiaire d'une chromatographie en phase liquide (spectre UV à 215 nm) (**B**) couplée à un spectromètre de masse (**C**).

Le dipeptide (**Figure 20A**) est détectable à une longueur d'onde de 215 nm, la molécule est éluée entre 1 et 1,5 min (**Figure 20B**).

La répartition isotopique ainsi que l'exactitude de mesure théorique prenant en compte l'abondance naturelle des atomes de carbone (C), hydrogène (H), azote (N), et oxygène (O) pour le dipeptide (C₉H₁₆N₂O₃) sont représentées par les rectangles mauves (**Figure 20C**). La mesure en spectrométrie de masse révèle un ion [M+H]⁺ à

201,1233 m/z, pour une valeur théorique similaire. Le profil isotopique à savoir 201,1233 m/z normalisé à 100 % présente un isotope à 202,1263 m/z d'une intensité relative de 9 %, et un dernier isotope à 203,1292 m/z d'une intensité de 0,6 % parfaitement corrélé aux valeurs théoriques attendues.

La comparaison entre l'analyse mesurée et l'analyse théorique de l'exactitude de mesure et de la répartition isotopique permet alors une identification formelle du dipeptide.

Afin d'évaluer la capacité de la synthétase AstS.Tr à assembler ce dipeptide, le surnageant d'une culture de B3R2 a été analysé selon la même méthode et comparé à celui issu de RIY261. Aucune production de dipeptide n'a pu être mise en évidence.

La fraction protéique (> 30 kDa) purifiée de ces deux souches a ensuite été utilisée pour réaliser l'assemblage *in vitro* du dipeptide (en présence d'ATP, de MgSO₄, et des deux monomères). Après 24 h de synthèse *in vitro* les peptides de taille inférieure à 10 kDa ont été analysés selon la méthode de dosage décrite précédemment. De nouveau, aucune production de dipeptide n'a pu être mise en évidence.

La protéine AstS.Tr, bien qu'étant traduite dans son ensemble, semble ne plus être fonctionnelle, probablement à la suite d'une scission entre la partie « 2 modules » et le domaine C_{term}.

4. Discussions

Le transfert de la voie de biosynthèse d'un peptide non-ribosomique de l'organisme le produisant naturellement à un hôte hétérologue est un processus complexe comportant plusieurs étapes. Le choix de l'hôte hétérologue, réceptacle des gènes préalablement identifiés comme étant impliqués dans la biosynthèse, est un paramètre crucial à considérer [6]. Le choix d'un organisme taxonomiquement proche limite le risque de générer un biais dans l'usage des codons qui entraînerait une diminution du taux de traduction, voire une absence de traduction par les ribosomes [44]. Les levures sont des hôtes hétérologues de premier choix pour exprimer des composés initialement produits par des champignons filamenteux. *S. cerevisiae*, par exemple, présente une bonne adaptabilité en termes d'usage des codons, ce qui lui confère une certaine versatilité pour l'expression de composés de manière hétérologue [45]. En revanche, les taux de production obtenus avec cet organisme sont souvent très faibles, voir plus faibles que chez le producteur homologue [21].

En ce qui concerne le transfert d'une voie de biosynthèse d'un composé de type NRP ou polykétide, le principal défi consiste à transférer une très large séquence. Le TAR cloning a permis au sein de cette étude d'intégrer le fragment de 17 kb codant pour l'astine synthétase au sein d'un plasmide d'expression. La séquence complète du gène requise pour la recombinaison est classiquement obtenue en co-transformant avec de l'ADNg ou ADNc rétro-transcrit, éventuellement préalablement restreint [24]. Malgré des prédispositions bien connues pour la recombinaison homologue, l'utilisation de cette méthode avec de l'ADNg restreint par *SgrDI* a montré ses limites. La longueur du gène *astS* génère une multitude de sites de restriction, rendant difficile le choix d'une enzyme capable de restreindre le génome sans altérer la séquence du gène. L'enzyme sélectionnée dans un premier temps (*SgrDI*) ne génère que trop peu de fragments, et de trop grandes tailles, ce qui restreint drastiquement l'efficacité du TAR cloning. Ce problème a pu être résolu en sélectionnant une enzyme de restriction (*EcoRI*) possédant de très nombreux sites de restriction au sein du génome de *C. asteris*, mais également un site de restriction au sein de *astS*. En réduisant le temps d'incubation de l'ADNg avec l'enzyme, son temps d'action peut être modulé, préservant ainsi (aléatoirement) une partie des sites de restriction qui auraient été coupés avec un temps d'incubation plus long. Cette approche, peu courante pour la préparation de l'ADN en vue d'un TAR cloning, nécessite une mise au point pour définir le temps d'incubation approprié pour générer des fragments d'ADNg.

La mise au point et le développement, depuis quelques années, d'ADN polymérase de plus en plus performantes, permet d'envisager l'amplification par PCR de la longue séquence à insérer. Le fragment obtenu via cette approche est beaucoup plus concentré et n'est pas inclus dans un enchevêtrement de fragments indésirables comme c'est le cas lors d'une restriction de l'ADN génomique. En utilisant une séquence amplifiée par PCR, les rendements en TAR cloning sont décuplés. Néanmoins, les ADN polymérase commercialisées, bien qu'étant capables d'amplifier des fragments de très longues tailles, sont susceptibles de générer des mutations au sein de la séquence amplifiée [46]. Pour remédier à ce problème, de nombreuses méthodes ont été

développées afin d'assembler de multiples fragments d'ADN en une longue séquence unique (Gibson assembly, USER, PCR overlap extension, etc.) [40–42]. L'amplification de fragments plus courts, assemblés par la suite, limite le risque de générer une mutation au sein de la séquence.

Le transfert du gène codant pour la synthétase au sein de l'hôte hétérologue ne garantit pas pour autant le succès de la production du NRP. De nombreuses étapes peuvent s'avérer limitantes dans la production d'une protéine hétérologue (transcription du gène étranger, compatibilité de codons permettant la traduction de la protéine, repliement correct de la protéine dans un environnement non naturel, etc.) [3]. Dans le cas d'une production hétérologue d'un NRP, certaines caractéristiques propres à la voie non-ribosomique complexifient l'opération. En effet, le processus de production hétérologue peut alors être divisé en deux phases : (1) l'expression, la traduction, et le maintien d'une macro-protéine NRPS hétérologue, (2) la conservation de la capacité de cette NRPS à assembler le NRP hétérologue. Cette deuxième étape implique notamment le repliement correct de la NRPS, son activation par le biais d'une phosphopantéthéine transférase (hétérologue dans le cas de *S. cerevisiae* et *Y. lipolytica*), ainsi que la disponibilité des précurseurs du NRP.

Les souches *S. cerevisiae* transformées avec le gène *astS* se sont avérées capables d'exprimer ce gène, que ce soit sous contrôle d'un promoteur inductible (GAL10) ou constitutif (TPI). En revanche, aucune traduction de la NRPS n'a pu être mise en évidence. Plusieurs facteurs ont été considérés pour tenter d'y remédier, comme l'utilisation de protocoles d'extractions adaptés aux protéines associées aux structures membranaires comme c'est le cas de certaines NRPSs [38,39]. Diverses méthodes de purification et concentration des protéines ont été mises en place afin de faciliter la détection de la NRPS. Un western blot, permettant de détecter jusqu'à des concentrations infimes de NRPS produites, a également été réalisé sans parvenir à obtenir de signal. Une production en quantité minimale, une instabilité de la NRPS, ou une dégradation de celle-ci sont des hypothèses pour expliquer l'absence de détection. Néanmoins, aux vues des analyses réalisées, il semble que dans les conditions considérées, *S. cerevisiae* ne soit pas capable de traduire cette enzyme pentamodulaire. La taille du complexe peut constituer un obstacle lors de la traduction, tout comme une incompatibilité des codons entre la levure et la séquence *astS*.

S. cerevisiae a été utilisée à plusieurs reprises avec succès pour l'expression hétérologue de synthétase de peptide non-ribosomique allant jusqu'à 3 modules. En particulier, la production hétérologue des NRPs d-(L- α -aminoadipyl)-L-cysteinyl-D-valine (ACV) [21,47] et fumiquinazoline F [48] a pu être accomplie chez cet hôte. Néanmoins, le transfert de la séquence codante n'est pas toujours un gage de succès en termes de production hétérologue du NRP. L'équipe d'Harvey *et al.* (2018) [49] a par exemple transféré 41 clusters de gènes issus d'organismes fongiques chez *S. cerevisiae*, et seuls 54 % des clusters clonés ont abouti à la production effective d'un composé hétérologue. La capacité de cette levure à exprimer une synthétase étrangère semble dépendante de la taille, de la nature de la synthétase, et probablement de l'organisme d'où elle provient. L'optimisation de la séquence génétique en termes d'usage des codons, préalablement au clonage, pourrait promouvoir largement l'étape

de traduction [50]. Mais cette opération constitue une approche encore onéreuse et sans réelle garantie de succès.

Au contraire de *S. cerevisiae*, la levure non-conventionnelle *Y. lipolytica* n'est pas un hôte courant pour l'expression hétérologue de synthétase non-ribosomique. Néanmoins, cette levure est capable de se multiplier jusqu'à des densités cellulaires très élevées et son potentiel de production de métabolites secondaires a déjà été mis en évidence notamment avec la production hétérologue de divers terpénoïdes [11,12]. Afin d'évaluer la capacité de *Y. lipolytica* à exprimer et exploiter une voie de synthèse non-ribosomique, seul un fragment du gène codant pour la synthétase d'astine (2 modules suivis du domaine C_{term}) a été cloné. La stratégie consistant à déplacer le domaine terminal plus en amont de la synthétase a déjà été explorée avec des domaines Te chez des NRPSs d'origine bactérienne [51]. Cette approche impliquant un déplacement de module / domaine permet notamment de générer de nouveaux composés NRPs qui ne sont pas assemblés naturellement par la synthétase [52]. La même stratégie a été mise en œuvre ici, en la transposant à un domaine terminal de type C_{term}, typique des NRPSs d'origine fongique.

La transcription du gène hybride intégré dans le génome de *Y. lipolytica* a pu être confirmée, tout comme la présence d'une protéine hétérologue. Néanmoins, les analyses par spectrométrie de masse et le western blot ont mis en évidence une séparation entre la partie protéique correspondant aux deux premiers modules et de celle du domaine C_{term}. En co-incubant ces deux fragments en présence des monomères nécessaires à l'assemblage du dipeptide attendu (L-2Abu – L-Pro), aucun composé n'a été généré. Dans le cas où le dipeptide assemblé comporte un acide aminé N-alkylé comme la proline, le NRP synthétisé peut être libéré sous une forme cyclique alternative en tant que diketopiperazine (DKP) [53]. La libération du dipeptide sous la forme 2Abu-Pro-DKP pourrait expliquer l'absence de détection du dipeptide linéaire. Néanmoins, il est plus probable que la scission du complexe NRPS ait entraîné une perte d'activité des différents domaines enzymatiques dans les conditions considérées. La présence d'une protéase de *Y. lipolytica* capable de reconnaître spécifiquement un site de clivage entre ces deux structures constitue l'hypothèse la plus probable pour expliquer cette scission. Les protéases de type dipeptidyl aminopeptidase sont un exemple de famille de protéases capables de générer une telle rupture [54]. Il est également possible que l'assemblage de ces deux fragments, non subséquents dans la NRPS native, ait généré une zone d'instabilité vulnérable. D'autant plus que la zone inter-domaine des NRPSs, aussi appelé « linker », est une zone très peu structurée, pauvre en feuilletés bêta et hélices alpha [55,56]. L'importance de ces linkers a par ailleurs récemment été démontrée dans la structuration globale et dans la spécificité de substrats des NRPS [57]. Le déplacement du domaine C_{term} en amont de la synthétase a pu engendrer une altération de l'un des linkers compromettant la stabilité structurale de la NRPS.

Que ce soit chez *S. cerevisiae* ou *Y. lipolytica*, l'expression hétérologue d'une voie non-ribosomique chez la levure nécessite un travail d'optimisation à plusieurs échelles. Les étapes de transformation et de transcription se sont avérées effectives chez chacun des transformants considérés, au contraire de l'étape de la traduction qui

constitue le principal obstacle. Le travail réalisé sur le modèle de l'astine a permis de mettre en lumière et de résoudre une partie des éléments à considérer pour aboutir à une production hétérologue de NRP. Néanmoins, malgré la production chez *Y. lipolytica* d'un fragment de la synthétase d'astine, aucun peptide non ribosomique hétérologue n'a pu être mis en évidence dans les conditions considérées.

5. Références

1. Grigoriev, I. V., et al. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.*, **42** (D1), D699–D704.
2. Weber, T., and Kim, H.U. (2016) The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.*, **1** (2), 69–79.
3. He, Y., et al. (2018) Recent advances in reconstructing microbial secondary metabolites biosynthesis in *Aspergillus* spp. *Biotechnol. Adv.*, **36** (3), 739–783.
4. Jahn, L. (2015) Characterization of a new endophytic astin producer, *Pelliciarosea asterica*, from *Aster tataricus*. *PhD Thesis*, (October 2015), 1–154.
5. Tsunematsu, Y., et al. (2013) Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.*, **9** (12), 818–825.
6. Anyaogu, D.C., and Mortensen, U.H. (2015) Heterologous production of fungal secondary metabolites in Aspergilli. *Front. Microbiol.*, **6** (FEB), 1–6.
7. Gidijala, L., et al. (2009) An engineered yeast efficiently secreting penicillin. *PLoS One*, **4** (12).
8. Schwecke, T., et al. (2006) Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. *ChemBioChem*, **7** (4), 612–622.
9. Erler, A., et al. (2006) Recombineering reagents for improved inducible expression and selection marker re-use in *Schizosaccharomyces pombe*. *Yeast*, **23** (11), 813–823.
10. Gao, L., et al. (2013) Engineered fungal polyketide biosynthesis in *Pichia pastoris*: a potential excellent host for polyketide production. *Microb. Cell Fact.*, **12** (1), 12–77.
11. Gao, S., et al. (2017) Iterative integration of multiple-copy pathway genes in *Yarrowia lipolytica* for heterologous β -carotene production. *Metab. Eng.*, **41**, 192–201.
12. Yang, X., et al. (2016) Heterologous production of α -farnesene in metabolically engineered strains of *Yarrowia lipolytica*. *Bioresour. Technol.*, **216**, 1040–1048.
13. Wagner, J.M., and Alper, H.S. (2015) Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances. *Fungal Genet. Biol.*, **89**, 126–136.
14. Kellner, H., et al. (2011) Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. *Microbiol. Res.*, **166** (5), 360–368.
15. Li, X., et al. (2013) A resurrected mammalian hAT transposable element and a closely related insect element are highly active in human cell culture. *Proc. Natl. Acad. Sci. U. S. A.*, **110** (6), E478–87.
16. Terpitz, U., et al. (2008) Electrofused giant protoplasts of *Saccharomyces cerevisiae* as a novel system for electrophysiological studies on membrane proteins. *Biochim. Biophys. Acta - Biomembr.*, **1778** (6), 1493–1500.
17. Barth, G., and Gaillardin, C. (1996) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*: *Biotechnological Applications.*, **19** (4), 219–237.
18. Carly, F., et al. (2018) Identification and characterization of EYD1, encoding an erythritol dehydrogenase in *Yarrowia lipolytica* and its application to bioconvert erythritol into erythrose. *Bioresour. Technol.*, **247**, 963–969.
19. Coutte, F., et al. (2010) Effect of *pps* disruption and constitutive expression of *srfA* on surfactin productivity, spreading and antagonistic properties of *Bacillus subtilis* 168 derivatives. *J. Appl. Microbiol.*, **109** (2), 480–491.
20. Raeder, U., and Broda, P. (1985) Rapid preparation of DNA from filamentous fungi. *Lett. Appl. Microbiol.*, **1** (1).

21. Siewers, V., et al. (2009) Heterologous production of non-ribosomal peptide LLD-ACV in *Saccharomyces cerevisiae*. *Metab. Eng.*, **11** (6), 391–397.
22. Trassaert, M., et al. (2017) New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. *Microb. Cell Fact.*, **16** (1), 141.
23. Fickers, P., et al. (2003) New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods*, **55** (3), 727–737.
24. Kouprina, N., and Larionov, V. (2016) Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma*, **125** (4), 621–632.
25. Gietz, R.D., and Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2** (1), 31–34.
26. Chen, D.C., et al. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.*, **48** (2), 232–235.
27. Ittisoponpisan, S., et al. (2019) Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? *J. Mol. Biol.*, **431** (11), 2197–2212.
28. Gao, A., et al. (2018) Evaluation of a Multiplex PCR for Detection of the Top Seven Shiga Toxin-Producing *Escherichia coli* Serogroups in Ready-to-Eat Meats, Fruits, and Vegetables. *J. AOAC Int.*, **101** (6), 1828–1832.
29. J. Livak, K., and Thomas D., S. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods.*, **25** (4), 402–408.
30. Nadai, C., et al. (2015) Selection and validation of reference genes for quantitative real-time PCR studies during *Saccharomyces cerevisiae* alcoholic fermentation in the presence of sulfite. *Int. J. Food Microbiol.*, **215**, 49–56.
31. Ma, B., et al. (2003) PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, **17** (20), 2337–2342.
32. Liang, J., et al. (2009) Effective elimination of nucleic acids from bacterial protein samples for optimized blue native polyacrylamide gel electrophoresis. *Electrophoresis*, **30** (14), 2454–2459.
33. Haynes, S.W., et al. (2013) Complexity generation in fungal peptidyl alkaloid biosynthesis: A two-enzyme pathway to the hexacyclic MDR export pump inhibitor ardeemin. *ACS Chem. Biol.*, **8** (4), 741–748.
34. Sharpl, P.M., and Li, W. (1987) potential applications Nucleic Acids Research. *Nucleic Acids Res.*, **15** (3), 1281–1295.
35. Jahn, L., et al. (2017) *Cyanoderrella asteris* sp. nov. (Ostropales) from the inflorescence axis of *Aster tataricus*. *Mycotaxon Issn*, **132** (March), 107–123.
36. Fan, L.-L., et al. (2019) Whole exome sequencing identified a novel mutation (p.Ala1884Pro) of β -spectrin in a Chinese family with hereditary spherocytosis. *J. Gene Med.*, **21** (2–3), e3073.
37. Martel, E., et al. (2019) Effect of foot-and-mouth disease virus 3C protease B 2 β -strand proline mutagenesis on expression and processing of the P1 polypeptide using a plasmid expression vector. *J. Gen. Virol.*, **100** (3), 446–456.
38. Imperi, F., and Visca, P. (2013) Subcellular localization of the pyoverdine biogenesis machinery of *Pseudomonas aeruginosa*: A membrane-associated “siderosome.” *FEBS Lett.*, **587** (21), 3387–3391.
39. Hantash, F.M., and Earhart, C.F. (2000) Membrane association of the *Escherichia coli* enterobactin synthase proteins EntB/G, EntE, and EntF. *J. Bacteriol.*, **182** (6), 1768–1773.
40. Bond, C., et al. (2016) *Saccharomyces cerevisiae* as a tool for mining, studying and engineering fungal polyketide synthases. *Fungal Genet. Biol.*, **89**, 52–61.
41. Cavaleiro, A.M., et al. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.*, **4** (9), 1042–1046.

42. Gibson, D.G., et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6** (5), 343–345.
43. Williams, S.A. (2001) SfiI endonuclease activity is strongly influenced by the non-specific sequence in the middle of its recognition site. *Nucleic Acids Res.*, **29** (7), 1476–1483.
44. Beites, T., and Mendes, M. V. (2015) Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial secondary metabolism. *Front. Microbiol.*, **6** (SEP), 1–10.
45. Tsunematsu, Y., et al. (2013) Yeast-based genome mining, production and mechanistic studies of the biosynthesis of fungal polyketide and peptide natural products. *Nat. Prod. Rep.*, **30** (8), 1139–1149.
46. Jia, H., et al. (2014) Long-range PCR in next-generation sequencing: Comparison of six enzymes and evaluation on the MiSeq sequencer. *Sci. Rep.*, **4** (5737).
47. Awan, A.R., et al. (2017) Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nat. Commun.*, **8** (May), 1–8.
48. Ishiuchi, K., et al. (2012) Establishing a New Methodology for Genome Mining and Biosynthesis of Polyketides and Peptides through Yeast Molecular Genetics. *ChemBioChem*, **13** (6), 846–854.
49. Harvey, C.J.B., et al. (2018) HEx: A heterologous expression platform for the discovery of fungal natural products. *Sci. Adv.*, **4** (4), eaar5459.
50. Ongley, S.E., et al. (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.*, **30** (8), 1121–1138.
51. Gao, L., et al. (2016) Translocation of the thioesterase domain for the redesign of plipastatin synthetase. *Sci. Rep.*, **6**, 1–9.
52. Calcott, M.J., and Ackerley, D.F. (2014) Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. *Biotechnol. Lett.*, **36** (12), 2407–2416.
53. Schwarzer, D., et al. (2001) Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chem. Biol.*, **8** (10), 997–1010.
54. Matoba, S., and Ogrzydziak, D.M. (1989) A novel location for dipeptidyl aminopeptidase processing sites in the alkaline extracellular protease of *Yarrowia lipolytica*. *J. Biol. Chem.*, **264** (11), 6037–6043.
55. Samel, S.A., et al. (2007) Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure*, **15** (7), 781–792.
56. Tarry, M.J., et al. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*, **25** (5), 783-793.
57. Farag, S., et al. (2019) Inter-Modular Linkers play a crucial role in governing the biosynthesis of non-ribosomal peptides. *Bioinformatics.*, btz127.

Chapitre 5

Discussion générale et Perspectives

1. Discussion générale

Cyanodermella asteris a été isolé très récemment (2015) à partir des racines d'*Aster tataricus* [1]. À ce jour, très peu d'informations sont disponibles concernant ce champignon filamenteux nouvellement identifié. Seule la voie de biosynthèse de l'antraquinone skyrin, un composé anti-cancéreux, a été élucidée chez ce champignon [2]. Les travaux menés dans le cadre de cette étude, combinés aux observations de Schafhauser et al. (2019) [3], ont permis d'attribuer la production de l'astine C, un autre composé anti-tumoral, à *C. asteris* plutôt qu'à son hôte *A. tataricus* comme décrit initialement [4,5]. Cet élément ouvre de nouvelles perspectives en termes de production d'astine C, qui jusqu'ici était extraite des racines d'*A. tataricus* avec de faibles rendements (0.16 mg/g de racines) [6]. Néanmoins, *C. asteris* présente un taux de croissance faible, ce qui limite la production de l'astine C. Au cours de cette étude, deux approches ont été considérées pour augmenter les taux de production : (1) optimisation de la production homologue par *C. asteris*, (2) développement d'un système de production hétérologue chez la levure. Un bilan de la stratégie mise en place pour chacune de ces approches est proposé sur la **Figure 1**.

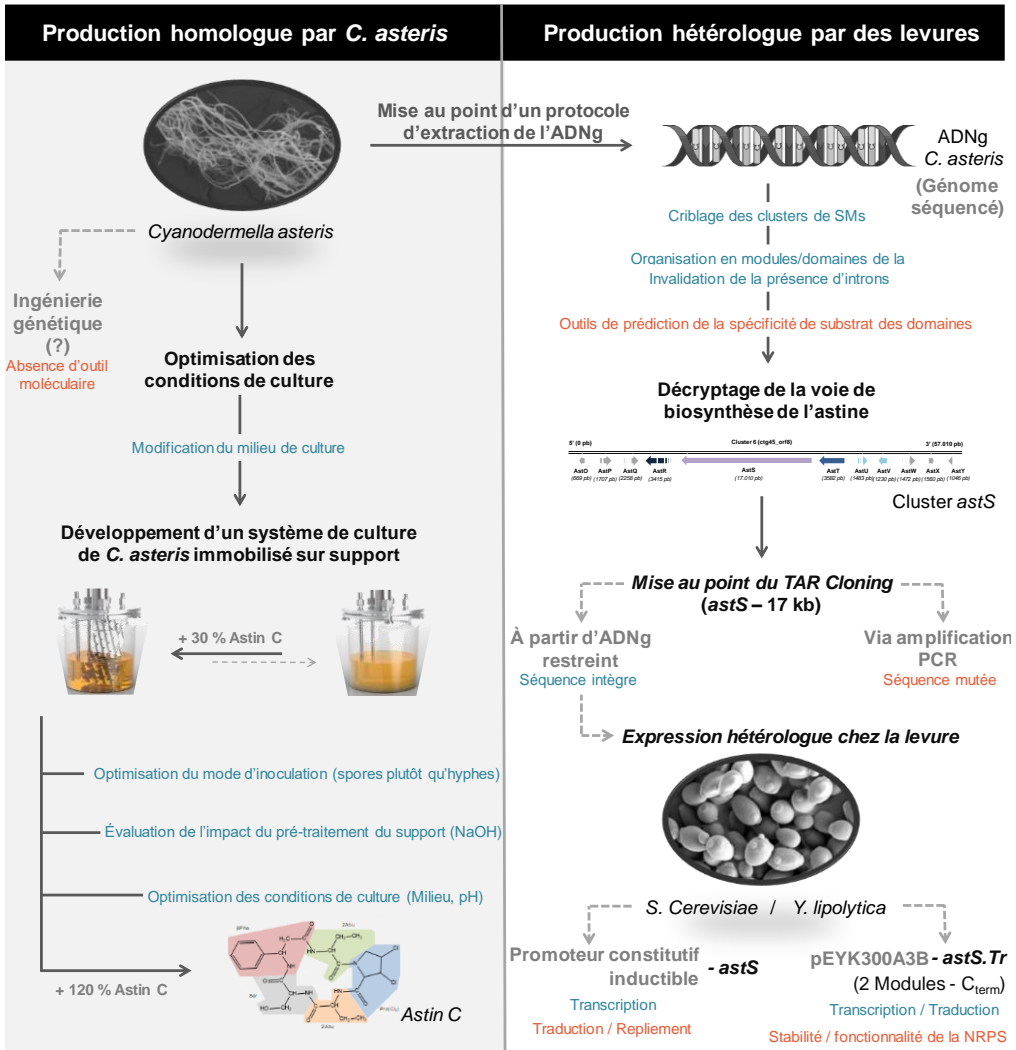


Figure 1 : Description des deux approches (homologue et hétérologue) développées au cours de cette étude afin d'accroître les rendements de production en astine C.

L'approche consistant à optimiser la production homologue chez *C. asteris* peut être divisée en deux options. La première option possible consiste à améliorer l'efficacité de la souche par ingénierie génétique. Par exemple, en positionnant les gènes impliqués dans la voie de biosynthèse sous contrôle d'un promoteur fort comme cela a déjà été réalisé avec le promoteur *PalcA* et les gènes NRPS *ivoA*, *ivoB*, et *ivoC*, chez *A. nidulans* [7]. Malheureusement, le manque d'informations concernant les promoteurs fonctionnels chez *C. asteris*, et l'absence totale d'outils moléculaires nécessaires à sa transformation, limitent l'utilisation de cette option.

Par conséquent, cette étude s'est focalisée sur l'optimisation des conditions de culture de ce champignon filamenteux afin d'accroître les taux de production d'astine

C. En premier lieu, la croissance et la productivité de *C. asteris* ont été évaluées dans des milieux de compositions différentes (dérivés d'un milieu à base d'extrait de malt). L'ajout dans le milieu de β -Phe, l'un des monomères composant l'astine C, a permis d'augmenter significativement les taux de production. L'impact positif de la supplémentation du milieu avec des précurseurs du NRP assemblé avait déjà été démontré, notamment chez *Verticillium lamellicola* avec une augmentation de la productivité en verlameline proportionnelle à la quantité d'acide 5-hydroxytetradécanoïque ajouté dans le milieu [8]. Dans le cas de la production d'astine, seul le précurseur β -Phe a eu un impact positif, puisque l'ajout de Ser et de Pro n'a pas permis d'accroître la production, et que la présence d'acide aminobutyrique a inhibé partiellement le développement du champignon. En se basant sur ce milieu optimisé, deux systèmes de cultures différents ont été évalués : la culture planctonique et la culture où les cellules fongiques sont immobilisées sur un support métallique. Ce second système a déjà été utilisé à de multiples reprises pour améliorer des rendements de production de métabolites secondaires chez divers organismes fongiques comme par exemple *Aspergillus terreus* [9], *Aspergillus oryzae* [10], *Aspergillus niger* [11], ou *Yarrowia lipolytica* [12]. Cette approche a été retenue car elle permet de reproduire partiellement les conditions de culture naturelles de *C. asteris*, se développant sur un support constitué des racines d'*A. tataricus* [1]. Ce mode de culture est donc susceptible d'être favorable au développement de ce champignon filamenteux.

En immobilisant *C. asteris* sur un support, la production d'astine a pu être augmentée de 30 % par rapport à un mode de culture en condition planctonique. Cette augmentation de la production est essentiellement corrélée avec une augmentation de la biomasse fongique de 39 % dans la condition sur support. Le champignon présente des morphologies différentes selon le mode de culture. En présence d'un support, il forme un biofilm dense et complexe, tandis qu'en culture planctonique il adopte une conformation sous forme de pelotes plus relâchées. De même, le réseau de polymères extracellulaires mis en place au niveau du biofilm est également beaucoup plus complexe et développé que celui visible au sein des pelotes fongiques. Le métabolisme du champignon a également été impacté avec une consommation plus lente des sucres présents dans le milieu. Ce phénomène peut s'expliquer par un accès beaucoup plus restreint au substrat au sein des biofilms par rapport aux cultures planctoniques [13]. L'impact de plusieurs paramètres de culture a ensuite été évalué afin d'optimiser la culture de ce champignon de manière immobilisée. Le pH du milieu de culture s'est avéré être un facteur déterminant à la fois pour la production d'astine C et le développement de *C. asteris*. Ces deux éléments ont été très largement favorisés par l'acidification du milieu induite naturellement par le champignon filamenteux. À l'inverse le maintien du pH à sa valeur initiale (4.8) a considérablement inhibé la croissance du champignon, et par conséquent la production d'astine C. Parmi les autres paramètres considérés, le mode d'innoculation du champignon via des spores plutôt qu'à partir d'hyphe permet de promouvoir l'attachement de la biomasse fongique sur le support. Bien qu'aucune information ne soit disponible concernant la composition des spores de *C. asteris*, la paroi des spores

des champignons filamenteux comporte souvent des fibrilles d'hydrophobines capables d'affecter les propriétés de surface et de promouvoir la fixation des spores [14,15]. La présence de ce type de composés sur les spores de *C. asteris* pourrait expliquer l'attachement préférentiel des spores, plutôt que des hyphes, sur le support métallique. En considérant les différents facteurs mis en évidence comme ayant un impact sur la croissance de *C. asteris* et sa production d'astine, des conditions de cultures optimisées ont été définies. Celles-ci consistent à cultiver *C. asteris*, inoculé à partir de spores, sur un support métallique préalablement traité au NaOH, immergé dans un milieu MEAc (dérivé du milieu malt extract MEA) où le pH n'est pas contrôlé. Ces conditions optimisées ont permis d'augmenter la production en astine C de 120 % par rapport aux conditions de cultures initiales. Le rendement spécifique d'astine obtenu dans ces conditions s'élève à 0,89 mg / g de biomasse fongique, ce qui correspond aux rendements d'astine obtenus lors d'une extraction réalisée à partir de 6 g de racine d'*A. tataricus* [6]. De plus, la production par fermentation de *C. asteris* permet de faciliter la récupération du métabolite synthétisé par rapport au mode de production à partir de la plante. Dans le cas du champignon filamenteux, l'astine C sécrétée dans le milieu de culture peut être extraite directement après avoir retiré le support sur lequel sont immobilisées les cellules. À l'inverse dans le cas de la plante, l'extraction débute par une étape contraignante de broyage des racines, suivie par de nombreuses étapes de purification. Bien que les taux de production de l'astine C à partir de *C. asteris* aient été plus que doublés au cours de cette étude, ils demeurent nettement insuffisants pour envisager une production à l'échelle industrielle.

Les taux de production n'ayant pas atteint les niveaux espérés, une seconde approche consistant à mettre en place une production hétérologue d'astine a été développée. Le principe consiste à transférer la voie de biosynthèse de l'astine chez un autre organisme dont les conditions de culture sont maîtrisées et adaptées à une production à l'échelle industrielle [16]. Afin de mettre en place cette approche, les différents gènes impliqués dans la voie de biosynthèse doivent préalablement être identifiés dans le génome de l'organisme produisant naturellement le composé d'intérêt. Une analyse bioinformatique a donc été réalisée sur le génome de *C. asteris* afin d'évaluer et de décrypter la voie de biosynthèse de l'astine. Les nombreux outils bioinformatiques développés ces dernières années (antiSMASH, InterProScan, PKS/NRPS Analysis, NaPDoS, NRPS Predictor2, etc.) facilitent l'analyse des génomes et l'identification de voies impliquées dans la production de métabolites secondaires telle que la voie non-ribosomique [17]. L'utilisation d'antiSMASH a permis de mettre en évidence, parmi les 38 clusters de gènes impliqués dans la biosynthèse de métabolites secondaires, un unique gène codant pour une NRPS à 5 modules capable d'assembler un pentapeptide tel que l'astine. Le cluster contenant ce gène a ensuite été analysé selon une méthodologie adaptée à partir du workflow FLORINE [18]. Afin de prendre en compte les spécificités relatives aux NRPSs d'origine fongique, un workflow alternatif a été mis en place pour l'analyse de *C. asteris*. Celui-ci comprend, en plus de l'utilisation des logiciels traditionnels, une évaluation de la présence d'introns, ainsi que le développement d'une méthode alternative de prédiction de la spécificité des domaines de la NRPS. Plusieurs

éléments importants ont pu être mis en évidence en appliquant ce workflow. Tout d'abord, l'organisation en domaine de la synthétase a été élucidée : A-T-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C_{term}. Une fois l'organisation en domaine déterminée, la position des deux introns issus de l'annotation du génome de *C. asteris* a été évaluée. Ces « introns » sont positionnés au sein des domaines A₁ et A₃. L'épissage des séquences correspondantes sur l'ARNm entraînerait par la suite la perte d'une des séquences protéiques conservées chez toutes les NRPSs (signatures). La présence d'introns au sein du gène codant pour cette NRPS a donc été invalidée. Cette approche constitue une méthode rapide pour évaluer la présence réelle d'un intron initialement annoté, mais elle n'est applicable que dans le cas où l'intron est localisé au sein d'une séquence codant pour une des signatures NRPS. La présence de 5 domaines d'adénylation permet l'assemblage linéaire d'un pentapeptide, qui est probablement cyclisé par le C_{term} lors de la libération du NRP. Le domaine C_{term}, spécifique des NRPSs fongiques, présente une activité cyclisatrice qui a été démontrée par exemple chez les NRPSs de la cyclosporine, la cyclochlorotine ou la tryptoquialanine [19–21]. L'analyse des domaines C par NaPDoS a permis de confirmer la proximité structurale du domaine C_{term} avec les domaines C_{term} de la cyclosporine (chez *Cochliobolus carbonum*) et de la cyclochlorotine (chez *T. islandicus*). En ce qui concerne l'identification du sous-type des différents domaines C de la NRPS, NaPDoS les a détectés en tant que sous-type « Dual C/E » (générant une épimérisation du monomère en plus de la réaction de condensation), ce qui n'est pas en accord avec ce qui est attendu pour la synthétase d'astine. En effet, seuls des monomères sous la forme L étant incorporés dans le peptide d'astine, les domaines C sont de sous type ^LC_L. Les différents domaines C présents dans l'arbre phylogénétique proposé par NaPDoS (basé sur la proximité des séquences protéiques) sont tous identifiés en tant que domaines « Dual ». Cependant, certains de ces domaines phylogénétiquement proches, comme par exemple le domaine C₂ de la synthétase de l'HC-toxine, sont en réalité des domaines catalysant la liaison peptidique entre deux L-acides aminés (sous-type ^LC_L) [21,22]. Il semble donc que les résultats fournis par ce logiciel pour l'analyse des domaines d'origine fongique soit biaisés. De même, les prédictions fournies par NRPS Predictor2 n'ont pas permis de mettre en évidence une spécificité des domaines A analysés pour les monomères incorporés dans l'astine (niveau de prédiction « nearest neighbor »). Néanmoins le substrat prédit pour le domaine A₄ (L-Cys) présente une structure très proche d'un des monomères de l'astine (L-Ser), et L-2Abu a bien été proposé parmi les substrats possibles pour les domaines A₁ et A₃ (niveau de prédiction « large cluster »). De manière générale, les prédictions réalisées sur les NRPSs sont souvent correctes concernant les NRPSs d'origine bactérienne, et très peu fiables pour les NRPSs fongiques [16]. L'analyse des séquences par l'intermédiaire des outils bioinformatiques se base généralement sur des comparaisons et des alignements avec des séquences élucidées repertoriées dans les bases de données. Les logiciels développés ces dernières années se basent donc sur l'ensemble des séquences comprises dans les bases de données sans prendre en compte les spécificités des NRPSs fongiques. Les séquences bactériennes étant largement majoritaires au sein de ces bases de données, les analyses réalisées sur les séquences fongiques sont souvent

biaisées. Une approche alternative pour la prédiction de la spécificité en substrat des domaines d'adénylation a été développée au cours de cette étude. L'idée consistait à modéliser la structure tridimensionnelle des différents modules de la NRPS à analyser, afin de réaliser des analyses en se basant sur ce modèle. En théorie, en évaluant la conformation de la poche catalytique des domaines d'adénylation, on peut estimer le type de substrat possiblement reconnu par ce site. Cette méthode a été validée sur un domaine A cristallisé (1AMU) dont la position du substrat a été expérimentalement validée. En considérant uniquement 1AMU (où le substrat Phe cristallisé a été retiré), les scores de docking ont permis d'attribuer de nouveau le résidu Phe comme substrat le plus probable. La même méthode a été transposée au module de la NRPS modélisée, en estimant la position du co-facteur AMP en fonction de sa position réelle dans 1AMU. Les résultats ont montré un positionnement relatif de l'AMP et des monomères testés (monomères incorporés dans l'astine ou issus des prédictions de NRPS predictor 2) ne permettant pas d'établir la réaction d'adénylation. L'approche de docking n'a pas permis ici d'obtenir des prédictions fiables, notamment à cause du positionnement approximatif de l'AMP, et de la proximité en termes de propriétés stéréochimiques des différents substrats évalués.

Malgré les prédictions approximatives obtenues lors de l'analyse de cette synthétase, de nombreux éléments ont tout de même permis de confirmer son implication dans la biosynthèse de l'astine. En premier lieu, il s'agit du seul gène codant pour une synthétase à 5 modules, s'achevant par un domaine C_{term} capable d'assembler un pentapeptide cyclique. Deuxièmement, cette synthétase présente une homologie de séquence protéique de 62 % avec la NRPS de la cyclochlorotine, un pentapeptide cyclique dont seul un monomère diffère de l'astine C [20]. Enfin, le gène NRPS est positionné dans un cluster au sein duquel on retrouve des gènes codant pour une amino-transférase et une phenylalanine aminomutase, potentiellement impliqués dans la biosynthèse des précurseurs de l'astine (L-2Abu et β -Phe).

Les 17 kb de la séquence codante pour l'astine synthétase (*astS*) étant dépourvus d'introns, il n'est pas nécessaire d'extraire les ARNm de *C. asteris* pour générer un ADN complémentaire dépourvu d'introns. Cet élément a son importance lorsque l'on sait que la transcription inverse de fragment de longue taille (> 10 kb) constitue une étape délicate susceptible de générer des mutations dans la séquence de l'ADNc obtenu [23]. Dans le cas de l'*astS*, seul l'ADN a donc été considéré pour la transformation. *C. asteris* étant un champignon nouvellement identifié, aucune méthode d'extraction n'a jusqu'ici été décrite. La mise au point d'un protocole d'extraction dédié a donc été nécessaire pour obtenir l'ADN génomique de ce champignon. Les cellules des champignons filamenteux possèdent des parois qui sont souvent très résistantes aux différentes méthodes de lyse. L'efficacité des différents protocoles dépend de la composition des parois cellulaires de l'organisme à lyser. Une extraction au phénol-chloroforme en présence de SDS permet par exemple d'extraire l'ADN_g des pathogènes de plantes *Fusarium oxysporum* et *Pyrenochaeta terrestris* [24], alors qu'elle est inefficace pour *C. asteris*. L'utilisation de l'azote liquide s'avère être la méthode la plus efficace pour lyser les cellules de *C. asteris*. Cette méthode est également celle préconisée pour la lyse d'autres champignons filamenteux comme

Beauveria bassiana, *Hirsutella citriformis*, *Isaria javanica*, et *Metarhizium anisopliae* [25]. Néanmoins, pour ces quatre derniers organismes, la lyse mécanique en présence de billes a permis d'obtenir de l'ADN génomique, ce qui n'est pas le cas pour *C. asteris* qui est résistant à cette approche.

Afin d'exprimer de manière hétérologue les séquences provenant de *C. asteris*, deux levures ont été sélectionnées : *S. cerevisiae* et *Y. lipolytica*. La première a été choisie pour sa capacité à réaliser une recombinaison homologue avec des rendements élevés facilitant ainsi le clonage de longues séquences d'ADN par TAR cloning [26]. De plus, l'aptitude de cette levure à intégrer et exploiter une voie de biosynthèse non-ribosomique hétérologue a été démontrée à plusieurs reprises [27–29]. La seconde levure a été sélectionnée pour sa capacité à se développer à de hautes concentrations cellulaires et à son potentiel élevé en matière de production de métabolites secondaires [30,31]. D'autres hôtes peuvent être considérés pour réaliser une expression hétérologue comme par exemple les souches de *Sreptomyces* qui sont celles les plus utilisées en matière de production hétérologue de NRPs [32]. Mais le choix d'un hôte taxonomiquement proche de l'organisme, d'où proviennent les séquences est préférable pour limiter le risque de générer un biais dans l'usage des codons qui impacterait négativement la traduction de la protéine hétérologue [33,34]. L'utilisation de souches de levures plutôt que de souches bactériennes est donc préférable pour l'expression des séquences provenant de *C. asteris*. La séquence de 17 kb codant pour l'astine synthétase a, dans un premier temps, été insérée dans un plasmide permettant l'expression dans *S. cerevisiae*. Deux approches ont été considérées pour le clonage de ce fragment par TAR cloning. La première impliquant une amplification par PCR permet d'obtenir un fragment pur avec des concentrations élevées. Bien que les taux de transformation chez *S. cerevisiae* soient très élevés en utilisant cette approche, le risque d'introduire des mutations lors de l'amplification est très important. Alternativement, le fragment de 17 kb peut être inséré dans le plasmide par recombinaison homologue avec de l'ADN génomique de *C. asteris* restreint [26]. Les taux de transformation observés en utilisant cette approche sont beaucoup plus bas, mais la séquence intégrée dans le plasmide conserve son intégrité. L'efficacité de la transformation peut directement être corrélée avec la qualité de la restriction de l'ADN génomique, matérialisée par le nombre et la taille des fragments générés lors de la restriction. L'insertion correcte de la séquence codante au sein de l'hôte hétérologue est la première étape cruciale, mais ne suffit pas forcément à obtenir une production hétérologue du NRP final. De nombreuses étapes par la suite peuvent s'avérer limitantes pour la production hétérologue. Tout d'abord, la transcription de la séquence hétérologue peut poser des problèmes si le promoteur n'est pas adapté ou si des facteurs de transcriptions sont manquants. Des études ont également montré qu'il existe des degrés de compatibilité promoteur-gène et qu'il est donc préférable d'évaluer plusieurs promoteurs pour l'expression d'un gène hétérologue [35]. Une fois le transcrit obtenu, la traduction de celui-ci en protéine peut constituer un obstacle si l'usage des codons n'est pas compatible entre la souche native et la souche hétérologue, ou si les introns ne sont pas reconnus correctement [36]. Par la suite, la NRPS fonctionnelle est obtenue uniquement si elle est correctement repliée, et

préservée de l'activité des protéases de l'hôte hétérologue. L'assemblage du NRP est ensuite possible à condition qu'un bras phosphopantéthéine soit ajouté sur les domaines de thiolation de la NRPS [37], et que les monomères précurseurs du peptide soit présents au sein de l'hôte hétérologue [38,39]. À la suite de la transformation chez *S. cerevisiae*, la présence du transcrit a pu être confirmée avec des niveaux similaires pour deux promoteurs différents, l'un constitutif (P_{TPI}) et l'autre inductible ($P_{GAL1/GAL10}$). En revanche, la NRPS correspondante n'a pas été détectée. Il est possible que l'usage des codons chez *S. cerevisiae* ne soit pas compatible avec les séquences provenant de *C. asteris*. L'analyse de l'adaptabilité des codons chez *S. cerevisiae* n'avait pourtant pas mis en évidence la présence de codons rares ou absents, compromettant la traduction du transcrit de *astS*. La dégradation de la protéine hétérologue par l'une des protéases de *S. cerevisiae*, ou même une taille trop conséquente du transcrit à traduire, constituent deux hypothèses éventuelles.

Afin d'exclure cette dernière hypothèse, seul un fragment du gène *astS*, codant pour les deux premiers modules de la NRPS suivi par le domaine C_{term} délocalisé, a été cloné chez *Y. lipolytica*. La délocalisation du domaine terminal d'une NRPS plus en amont de la synthétase est une approche qui a déjà été utilisée chez la NRPS de la tyrocidine. Le domaine Te a dans ce cas-ci été déplacé à la suite du module 2 (C-A-T) de TycB permettant de générer un dipeptide D-Phe – L-Pro [40]. Alternativement, de nouveaux NRPs, non naturellement produits, peuvent également être générés en échangeant les unités modulaires (C-A-T) au sein de la NRPS ou entre différentes NRPSs [41]. L'approche développée chez *Y. lipolytica* au cours de cette étude est la première à mettre en œuvre le déplacement d'un domaine C_{term} au sein d'une NRPS. La construction réalisée a pour but avant tout d'évaluer la capacité de cette levure à exprimer et exploiter une voie non-ribosomique, mais également d'étudier la possibilité de générer de nouveaux NRPs en déplaçant le domaine C_{term} . Cette séquence *astS* tronquée, intégrée dans le génome de *Y. lipolytica*, a été correctement transcrite et traduite par la levure. Néanmoins, le complexe présente une instabilité et une zone de rupture a été mise en évidence entre la partie « 2 modules » et la partie C_{term} . Comme chez *S. cerevisiae*, l'action de protéases de l'hôte hétérologue pourrait expliquer cette zone de coupure spécifique. D'autant plus que cette zone inter-domaine, appelée « linker », est une zone très peu structurée avec peu d'hélice alpha et feuillet beta, ce qui la rend vulnérable aux protéases [42,43]. Aucune production de dipeptide (L-2Abu – L-Pro) n'a été détectée, confirmant l'absence de fonctionnalité de ce complexe. Abstraction faite de la rupture entre les deux fragments, le maintien de l'activité du domaine C_{term} n'était pas assuré puisque les domaines de condensation présentent une certaine spécificité de substrat qui, si elle n'est pas respectée, peut compromettre l'activité du domaine dans une synthétase hybride [44]. Le domaine C_{term} de l'AstS présente a priori une spécificité de substrat pour la β -Phe (monomère reconnu par le dernier module de l'AstS), et non pas pour la L-Pro (monomère reconnu par le module 2). Néanmoins, aucune information n'est disponible concernant la spécificité des domaines C_{term} , puisqu'elle a été décrite uniquement chez les domaines de condensation interne, où elle est plus marquée sur la partie « receveuse » du sous-domaine C [45].

L'approche hétérologue développée au cours de cette étude n'a pas permis d'aboutir à la production d'un NRP. Chez *S. cerevisiae* et *Y. lipolytica*, la transcription des séquences NRPSs a été confirmée, mais aucune protéine traduite n'a été détectée chez la première, et la NRPS synthétisée chez la seconde s'est avérée instable. Le clonage de la séquence d'intérêt n'est pas un gage de succès pour l'obtention du NRP final. Les mécanismes mis en jeu lors d'une production hétérologue, et d'autant plus de NRP, ne sont pas encore totalement maîtrisés. Les outils génétiques disponibles actuellement permettent de surmonter les obstacles en matière de clonage, voir même de transcription des séquences, mais il est très difficile d'influencer sur les étapes suivantes (traduction, repliement des protéines, maintien de la NRPS, etc). Cette étude sur le modèle de l'astine a mis en évidence un certain nombre de facteurs à prendre en compte pour la production hétérologue de NRPs. D'autres NRPSs ont, par le passé, été exprimées de manière hétérologue chez des levures comme les synthétases de l'ACV [28,39], la fumiquinazoline F [19,29], de la brevianamide F [27], et de la beauvericine [46]. Pour ces différents NRPs exprimés de manière non-ribosomique, le bilan en termes de productivité est mitigé. Dans le cas de brevianamide F, par exemple, la production hétérologue a permis de multiplier par 100 les taux de production en comparaison avec la production homologue chez *A. fumigatus* (de 0,383 à 35,6 mg/L). À l'inverse, dans le cas de l'ACVS (assemblant l'ACV précurseur de la pénicilline), les taux de production hétérologue chez *S. cerevisiae* ou *H. polymorpha* après optimisation atteignent à peine les productions obtenues chez la souche sauvage de *P. chrysogenum* [28,35,39]. Or cette souche sauvage a été optimisée pendant des décennies (au niveau de ses conditions de culture et de sa génétique), et les souches utilisées aujourd'hui dans l'industrie présentent des rendements de production au moins 500 fois supérieurs à la souche sauvage initiale [47]. En conclusion, il apparaît que l'approche hétérologue peut difficilement concurrencer les rendements obtenus avec une souche homologue d'un modèle bien établi et optimisé depuis plusieurs années. En revanche, elle peut être avantageuse dans le cas où les niveaux de production homologue du NRP sont très bas voir inexistantes.

Dans le cas de l'astine, une production hétérologue pourrait permettre de surpasser les taux de production obtenus chez *C. asteris*. Néanmoins, même en cas de succès de l'expression hétérologue d'*astS*, seule la forme non-chlorée de l'astine serait produite (forme G). Cette forme, au contraire de l'astine di-chlorée (forme C) ne présente aucune activité anti-tumorale. L'identification de l'halogénase ou du mécanisme impliqué dans la chloration de l'astine G chez *C. asteris*, puis le transfert des gènes chez l'hôte hétérologue, constituent donc de nouvelles étapes nécessaires à l'obtention de l'astine C. En comparaison, l'approche homologue propose déjà un système de production transposable à l'échelle industrielle. Bien qu'une optimisation des conditions de culture, voire de la souche de *C. asteris*, soit encore nécessaire pour augmenter les taux de production, cette approche constitue la piste la plus prometteuse pour la production d'astine C à court et moyen terme.

2. Perspectives

2.1. Relatives à l'approche de production homologue

Les taux de production de l'astine C à partir de *C. asteris* étant insuffisants pour envisager une production à l'échelle industrielle, une optimisation des conditions de cultures est encore nécessaire pour obtenir des rendements plus élevés. En particulier, la composition du milieu de culture peut encore être améliorée. En effet, la composition très riche de ce dernier, en plus de le rendre coûteux, interfère avec l'étape d'extraction de la molécule d'astine qui doit être purifiée à partir d'un mélange complexe d'éléments nutritifs. D'autre part, l'impact positif de l'acidification du milieu sur le développement du champignon peut également être exploité dans un système de culture où le pH serait maintenu à un niveau plus acide que celui testé ici (pH 4,8 acidifié jusqu'à 3,0 lors de la croissance de *C. asteris*).

Le système de culture du champignon en cellules immobilisées peut être intégré dans un mode de culture semi-continu qui est économiquement très intéressant. Dans un tel système, à l'issue de la première fermentation, le milieu de culture contenant l'astine C (dépourvu de cellules) peut être soutiré afin d'extraire le composé d'intérêt. Le support contenant les cellules de *C. asteris* déjà développées et immobilisées est, quant à lui, immergé de nouveau dans un milieu frais afin d'amorcer une seconde production. Un mode de culture similaire est par exemple utilisé pour produire des β -galactosidase et β -fructofuranosidase à partir respectivement de cellules immobilisées d'*Aspergillus oryzae* et d'*A. japonicus* [48,49]. Cependant, la mise en place d'un système de production semi-continu nécessite néanmoins plusieurs mises au point afin de déterminer le temps de culture idéal entre chaque remplacement de milieu, le nombre de cycles de culture réalisables à partir d'un même lot de cellules immobilisées, et éventuellement les variations des taux de production d'un cycle à un autre. En tenant compte de ces différents éléments la production d'astine C par *C. asteris* immobilisé pourrait être maximisée.

Alternativement, l'utilisation de méthodes induisant une mutagenèse aléatoire (rayons X, UV, traitement chimique) pourrait permettre de générer des mutants de *C. asteris* présentant potentiellement des taux de production plus élevés. Cette approche a été notamment utilisée au cours de décennies d'optimisation sur les souches *A. niger* et *P. chrysogenum* pour générer des mutants surproducteurs respectivement d'acide citrique et de pénicilline [47,50,51]. Parallèlement, le développement d'outils moléculaires adaptés pour la transformation de *C. asteris* ouvrirait de nouvelles perspectives en termes d'optimisation de la souche. Le positionnement des gènes impliqués dans la biosynthèse de l'astine (codant pour l'astine synthétase et la phenylalanine aminomutase par exemple) sous contrôle d'un promoteur fort permettrait de surexprimer ces gènes et d'augmenter la productivité. De même, l'ingénierie génétique pourrait être utilisée pour inactiver certains des 38 clusters de gènes identifiés comme étant impliqués dans la biosynthèse de métabolites secondaires chez *C. asteris*. Il est possible que l'une de ces voies métaboliques, encore non-élucidée, soit en compétition avec la voie de biosynthèse de l'astine concernant certains précurseurs. En particulier, la β -Phe, monomère limitant dans la production

d'astine, pourrait être monopolisée par la voie de biosynthèse d'un autre métabolite secondaire.

2.2. Relatives au décryptage des voies de biosynthèse des NRPs par analyses bioinformatiques à partir de génomes fongiques

Les prédictions réalisées à partir des séquences de *C. asteris* concernant la synthétase d'astine se sont illustrées par leur manque de fiabilité. Les problèmes relatifs à l'analyse des NRPSs d'origine fongique sont récurrents et soulignent la nécessité de mettre au point de nouveaux outils plus adaptés. Deux perspectives peuvent être explorées pour apporter une solution à cette problématique. La première consiste à mettre au point de nouveaux outils dédiés uniquement à l'analyse des séquences fongiques. La seconde consiste à adapter les outils existants pour que les analyses des séquences fongiques soient basées uniquement sur les séquences fongiques répertoriées dans les bases de données sans tenir compte des données d'origine bactérienne. Cette seconde solution nécessite l'approvisionnement au sein des bases de données d'un nombre plus important de NRPSs fongiques de référence, dont la spécificité des domaines A et le sous type des domaines C ont été validés expérimentalement.

En ce qui concerne la voie de biosynthèse de l'astine, malgré l'identification du gène codant pour la synthétase et du cluster associé, plusieurs gènes indispensables à la synthèse d'astine chez *C. asteris* n'ont pas encore été identifiés. C'est le cas par exemple du gène codant pour la phosphopantéthéine transférase nécessaire à l'activation de la NRPS d'astine, ou du gène codant pour une halogénase potentiellement impliquée dans la chlorination de la proline incorporée dans l'astine, permettant de générer les formes F et C. Par conséquent, des analyses bioinformatiques supplémentaires sont requises afin d'identifier ces deux gènes au sein du génome de *C. asteris*.

2.3. Relatives à la production hétérologue de l'astine

La stratégie développée au cours de cette étude n'a pas permis d'aboutir à la production d'astine par voie hétérologue. Chez *S. cerevisiae*, la production semble interrompue à l'étape de la traduction de la NRPS. Une optimisation de la séquence codante en termes de codons pourrait permettre de favoriser la traduction de l'ARNm. Néanmoins cette solution demeure couteuse et ne présente aucune garantie de succès. Chez *Y. lipolytica*, la traduction de la synthétase tronquée a pu être mise en évidence. La production semble cette fois limitée par une instabilité de la structure "2 modules - C_{term}". L'insertion de l'ensemble de la séquence codante pour *astS* pourrait permettre l'expression d'une protéine stable et potentiellement fonctionnelle. L'utilisation de la méthode du TAR cloning pour l'assemblage de longue séquence est plus délicat chez *Y. lipolytica* que chez *S. cerevisiae*. Néanmoins, d'autres méthodes permettent aujourd'hui d'assembler de larges séquences et d'envisager l'expression hétérologue

des 17 kb de *astS* chez cette levure (Gibson assembly, USER, PCR overlap extension, etc.) [52–54].

Une autre perspective serait d'envisager le transfert du gène *astS* chez un autre hôte hétérologue taxonomiquement plus proche de *C. asteris*. Les souches d'*Aspergillus* spp. utilisées à maintes reprises pour exprimer de manière hétérologue des synthétases issues d'autres champignons filamenteux, apparaissent comme des hôtes particulièrement adaptés [38,55–61]. Le clonage du gène *astS* est désormais facilité puisque celui-ci est isolé au sein du plasmide pESC-sfp-AstS, et peut donc facilement être libéré par une simple restriction enzymatique. Il est dorénavant aisé de le manipuler et de le transférer dans un hôte capable de l'exprimer et d'exploiter pleinement la voie de biosynthèse de l'astine.

3. Références

1. Jahn, L., et al. (2017) *Cyanodermeella asteris* sp. nov. (Ostropales) from the inflorescence axis of *Aster tataricus*. *Mycotaxon Issn*, **132** (March), 107–123.
2. Jahn, L., et al. (2017) Linking secondary metabolites to biosynthesis genes in the fungal endophyte *Cyanodermeella asteris*: The anti-cancer bisanthraquinone skyrin. *J. Biotechnol.*, **257** (December 2016), 233–239.
3. Thomas, S., et al. (2019) Production of antitumor astins by the endophyte *Cyanodermeella asteris* isolated from the medicinal plant *Aster tataricus*. *Submitted*.
4. Morita, H., et al. (1993) Astins A and B, antitumor cyclic pentapeptides from *Aster tataricus*. *Chem. Pharm. Bull. (Tokyo)*, **41** (5), 992–993.
5. Li, F., et al. (2018) Design and synthesis of plant cyclopeptide Astin C analogues and investigation of their immunosuppressive activity. *Bioorganic Med. Chem. Lett.*, **28** (14), 2523–2527.
6. Shen, Y., 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** (3), 260–268.
7. Sung, C.T., et al. (2017) Overexpression of a three-gene conidial pigment biosynthetic pathway in *Aspergillus nidulans* reveals the first NRPS known to acetylate tryptophan. *Fungal Genet. Biol.*, **101**, 1–6.
8. Ishidoh, K.I., et al. (2014) Identification of a gene cluster responsible for the biosynthesis of cyclic lipopeptide verlamelin. *Appl. Microbiol. Biotechnol.*, **98** (17), 7501–7510.
9. Baños, J.G., et al. (2009) High lovastatin production by *Aspergillus terreus* in solid-state fermentation on polyurethane foam: An artificial inert support. *J. Biosci. Bioeng.*, **108** (2), 105–110.
10. Zune, Q., et al. (2015) A fungal biofilm reactor based on metal structured packing improves the quality of a GlA :: GFP fusion protein produced by *Aspergillus oryzae*. 6241–6254.
11. Papagianni, M., and Matthey, M. (2004) Physiological aspects of free and immobilized *Aspergillus niger* cultures producing citric acid under various glucose concentrations. *Process Biochem.*, **39** (12), 1963–1970.
12. Vandermies, M., et al. (2018) *Yarrowia lipolytica* morphological mutant enables lasting in situ immobilization in bioreactor. *Appl. Microbiol. Biotechnol.*, **102** (13), 5473–5482.
13. Fanning, S., and Mitchell, A.P. (2012) Fungal biofilms. *PLoS Pathog.*, **8** (4), 1–4.
14. Beauvais, A., et al. (2014) *Aspergillus* cell wall and biofilm. *Mycopathologia*, **178** (5–6), 371–377.
15. Macindoe, I., et al. (2011) Self-assembly of functional, amphipathic amyloid monolayers by the fungal hydrophobin EAS. *PNAS.*, **109** (14), E804–E811.
16. Nielsen, J.C., and Nielsen, J. (2017) Development of fungal cell factories for the production of secondary metabolites: Linking genomics and metabolism. *Synth. Syst. Biotechnol.*, **2** (1), 5–12.
17. Weber, T., and Kim, H.U. (2016) The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth. Syst. Biotechnol.*, **1** (2), 69–79.
18. Caradec, T., et al. (2014) Prediction of Monomer Isomery in Florine: A Workflow Dedicated to Nonribosomal Peptide Discovery. *PLoS One*, **9** (1), e85667.
19. Xue Gao, et al. (2012) Cyclization of Fungal Nonribosomal Peptides by a Terminal Condensation-Like Domain. *Nat Chem Biol*, **412** (2), 426–434.

20. Schafhauser, T., et al. (2016) The cyclochlorotine mycotoxin is produced by the nonribosomal peptide synthetase CctN in *Talaromyces islandicus* ('*Penicillium islandicum*'). *Environ. Microbiol.*, **18** (11), 3728–3741.
21. Zhang, J., et al. (2016) Structural basis of nonribosomal peptide macrocyclization in fungi. *Nat. Chem. Biol.*, **12** (12), 1001–1003.
22. Weber, G., et al. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.*, **26** (2), 120–125.
23. Fu, G.K., et al. (2003) Full-length cDNA synthesis for long-distance RT-PCR of large mRNA transcripts. *Biotechniques.*, **34** (4), 768–770.
24. González-Mendoza, D., et al. (2010) A rapid method for isolation of total DNA from pathogenic filamentous plant fungi. *Genet. Mol. Res.*, **9** (1), 162–166.
25. Serna-Domínguez, M.G., et al. (2018) Two efficient methods for isolation of high-quality genomic DNA from entomopathogenic fungi. *J. Microbiol. Methods.*, **148**, 55–63.
26. Kouprina, N., and Larionov, V. (2016) Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma*, **125** (4), 621–632.
27. Tsunematsu, Y., et al. (2013) Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.*, **9** (12), 818–825.
28. Siewers, V., et al. (2009) Heterologous production of non-ribosomal peptide LLD-ACV in *Saccharomyces cerevisiae*. *Metab. Eng.*, **11** (6), 391–397.
29. Ishiuchi, K., et al. (2012) Establishing a New Methodology for Genome Mining and Biosynthesis of Polyketides and Peptides through Yeast Molecular Genetics. *ChemBioChem*, **13** (6), 846–854.
30. Gao, S., et al. (2017) Iterative integration of multiple-copy pathway genes in *Yarrowia lipolytica* for heterologous β -carotene production. *Metab. Eng.*, **41**, 192–201.
31. Yang, X., et al. (2016) Heterologous production of α -farnesene in metabolically engineered strains of *Yarrowia lipolytica*. *Bioresour. Technol.*, **216**, 1040–1048.
32. Nah, H.-J., et al. (2017) Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species. *Front. Microbiol.*, **8** (March).
33. Anyaogu, D.C., and Mortensen, U.H. (2015) Heterologous production of fungal secondary metabolites in *Aspergilli*. *Front. Microbiol.*, **6** (FEB), 1–6.
34. Beites, T., and Mendes, M. V. (2015) Chassis optimization as a cornerstone for the application of synthetic biology based strategies in microbial secondary metabolism. *Front. Microbiol.*, **6** (SEP), 1–10.
35. Awan, A.R., et al. (2017) Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nat. Commun.*, **8** (May), 1–8.
36. Ongley, S.E., et al. (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.*, **30** (8), 1121–1138.
37. Stachelhaus, T., et al. (1996) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem. Biol.*, **3** (11), 913–21.
38. Richter, L., et al. (2014) Engineering of *Aspergillus niger* for the production of secondary metabolites. *Fungal Biol. Biotechnol.*, **1** (1), 4.
39. Gidijala, L., et al. (2009) An engineered yeast efficiently secreting penicillin. *PLoS One*, **4** (12).
40. Schwarzer, D., et al. (2001) Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chem. Biol.*, **8** (10), 997–1010.
41. Duerfahrt, T., et al. (2003) Construction of hybrid peptide synthetases for the production of α -L-aspartyl-L-phenylalanine, a precursor for the high-intensity sweetener aspartame. *Eur. J. Biochem.*, **270** (22), 4555–4563.
42. Samel, S.A., et al. (2007) Structural and Functional Insights into a Peptide Bond-Forming

- Bidomain from a Nonribosomal Peptide Synthetase. *Structure*, **15** (7), 781–792.
43. Tarry, M.J., et al. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*, **25** (5), 783–793.e4.
 44. Bozhüyük, K.A.J., et al. (2019) Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.*, **11** (7), 653–661.
 45. Belshaw, P.J., et al. (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284** (5413), 486–489.
 46. Yu, D., et al. (2017) Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.*, **8** (May), 1–11.
 47. Jørgensen, H., et al. (1995) Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnol. Bioeng.*, **46** (2), 117–131.
 48. Guerrero, C., et al. (2015) Repeated-batch operation for the synthesis of lactulose with β -galactosidase immobilized by aggregation and crosslinking. *Bioresour. Technol.*, **190**, 122–131.
 49. Mussatto, S.I., et al. (2009) beta-Fructofuranosidase production by repeated batch fermentation with immobilized *Aspergillus japonicus*. *J. Ind. Microbiol. Biotechnol.*, **36** (7), 923–928.
 50. Gardner, J.F., et al. (2009) Production of Citric Acid by Mutants of *Aspergillus niger*. *J. Gen. Microbiol.*, **93** (2), 125–130.
 51. Newbert, R.W., et al. (1997) Analysis of a commercially improved *Penicillium chrysogenum* strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. *J. Ind. Microbiol. Biotechnol.*, **19** (1), 18–27.
 52. Bond, C., et al. (2016) *Saccharomyces cerevisiae* as a tool for mining, studying and engineering fungal polyketide synthases. *Fungal Genet. Biol.*, **89**, 52–61.
 53. Cavaleiro, A.M., et al. (2015) Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.*, **4** (9), 1042–1046.
 54. Gibson, D.G., et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6** (5), 343–345.
 55. Maiya, S., et al. (2006) The fumitremorgin gene cluster of *Aspergillus fumigatus*: Identification of a gene encoding brevianamide F synthetase. *ChemBioChem*, **7** (7), 1062–1069.
 56. Unkles, S.E., et al. (2014) Synthetic biology tools for bioprospecting of natural products in eukaryotes. *Chem. Biol.*, **21** (4), 502–508.
 57. Geib, E., and Brock, M. (2017) ATNT: an enhanced system for expression of polycistronic secondary metabolite gene clusters in *Aspergillus niger*. *Fungal Biol. Biotechnol.*, **4** (1), 13.
 58. Zobel, S., et al. (2016) Reprogramming the Biosynthesis of Cyclodepsipeptide Synthetases to Obtain New Enniatins and Beauvericins. *ChemBioChem*, **17** (4), 283–287.
 59. Brandenburger, E., et al. (2017) A Highly Conserved Basidiomycete Peptide Synthetase Produces a Trimeric Hydroxamate Siderophore. *Appl. Environ. Microbiol.*, **83** (21), 1–13.
 60. De Mattos-Shiple, K.M.J., et al. (2018) The cycloaspeptides: Uncovering a new model for methylated nonribosomal peptide biosynthesis. *Chem. Sci.*, **9** (17), 4109–4117.
 61. Yoshimi, A., et al. (2018) Heterologous production of a novel cyclic peptide compound, KK-1, in *Aspergillus oryzae*. *Front. Microbiol.*, **9** (APR), 1–12.

Communications scientifiques

Congrès scientifiques

- **Octobre 2017**

Communication Orale au congrès de la Société Française de Microbiologie (Paris).
« Heterologous production of NRPs: From the identification of the biosynthesis genes to the productive heterologous yeast ».

Vassaux A., Tarayre C., Steels S., Vandenbol M., Jacques P., Fickers P., Leclère V.

- **Janvier 2018**

Communication Poster au Symposium INTERREG (Gand)
« Heterologous production of NRPs: From the identification of the biosynthesis genes to the productive heterologous yeast ».

Vassaux A., Tarayre C., Steels S., Vandenbol M., Jacques P., Fickers P., Leclère V.

- **Septembre 2018**

Communication Orale au congrès international ESBES2018 (Lisbonne)
« Astin C production by the novel fungus *Cyanodermella asteris* in planktonic conditions in comparison to biofilm state ».

Vassaux A., Tarayre C., Théatre A., Arguëlles-Arias A., Telek S., Compère P., Vandenbol M., Leclère V., Fickers P., Delvigne F., Ongena M., Jahn L., Ludwig-Müller J., Van Berkel W. J. H., Schafhauser T., Wohlleben W., Van Pée K., Willems L., Jacques P.

Articles scientifiques

Vassaux, A., Tarayre, C., Arguëlles-Arias, A., Compère, P., Delvigne, F., Fickers, P., Jahn, L., Lang, A., Leclère, V., Ludwig-Müller, J., Ongena, M., Schafhauser, T., Telek, S., Théatre, A., Berkel, W.J.H., Vandenbol, M., Pée, K., Willems, L., Wohlleben, W., Jacques, P., **2019**. Astin C production by the endophytic fungus *Cyanodermella asteris* in planktonic and immobilized culture conditions. *Biotechnology Journal*. <https://doi.org/10.1002/biot.201800624>

Vassaux, A., Meunier, L., Vandenbol, M., Baurain, D., Fickers, P., Jacques, P., Leclère, V., **2019**. Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production. *Biotechnology Advances*. <https://doi.org/10.1016/j.biotechadv.2019.107449>