

## Résumé

Titre de la thèse :

**Les cyclodipeptide synthases : vers la compréhension de leur mécanisme catalytique et des bases moléculaires de leur spécificité**

Auteur : **Yan Li**

Ecole doctorale: **ED 425 (Innovation thérapeutique : du fondamental à l'appliqué)**

Année : **2009 – 2012**

Série doctorat N : **1185**

Les substances naturelles issues des microbes, des plantes ou encore des animaux ont de tout temps été utilisées par l'Homme pour leurs vertus thérapeutiques. La biodiversité du milieu naturel n'a pas encore trouvé d'égal pour la découverte de médicaments ; ainsi, les substances naturelles ou les molécules qui en dérivent directement représentent plus de 60% des médicaments mis sur le marché médicamenteux. Comprendre comment ces substances naturelles sont synthétisées dans l'organisme producteur revêt un intérêt scientifique considérable. En effet, cette compréhension donne accès à une amélioration de la production des substances naturelles mais aussi à la

possibilité de générer de nouvelles molécules mimant les substances naturelles d'intérêt et présentant des activités pharmacologiques améliorées.

Le travail de thèse présenté dans ce manuscrit porte sur l'étude de la biosynthèse des cyclodipeptides et de leurs dérivés complexes, les dicétopipérazines (DKP), qui constituent une classe de molécules naturelles extrêmement répandues chez les microorganismes. Les DKP se caractérisent par une grande variété de structures chimiques et une large gamme d'activités biologiques et pharmacologiques.

Dans l'introduction du manuscrit, je décris la découverte et la caractérisation préliminaire au laboratoire d'une famille d'enzymes, les synthases de cyclodipeptides (CDPS), qui sont impliquées dans la biosynthèse de certaines DKP (Gondry, Sauguet et al. 2009). Ce travail avait permis d'identifier huit CDPS dans des bactéries de phyla différents et de déterminer les cyclodipeptides produits par chacune de ces CDPS. Il avait également permis de montrer que les CDPS détournent des ARNt aminoacylés (aa-ARNt) de leur rôle essentiel dans la synthèse protéique ribosomale pour les utiliser comme substrats et catalyser la formation des deux liaisons peptidiques des différents cyclodipeptides synthétisés. De plus, ce travail avait établi que les CDPS travaillent de concert avec des enzymes de modification des cyclodipeptides dans des voies de biosynthèse dédiées à

la biosynthèse de DKP (Belin et al. 2009 ; Belin, Moutiez et al. 2012). Ce travail antérieur à mon arrivée dans le laboratoire étant présenté, je me suis ensuite attachée à présenter mon sujet de thèse portant sur la caractérisation plus approfondie des CDPS avec pour objectif d'apporter des informations sur les bases moléculaires de la reconnaissance entre les CDPS et leurs substrats et sur le mécanisme catalytique emprunté par ces enzymes. Le manuscrit est composé de six chapitres, le premier chapitre étant une étude bibliographique, les chapitres deux à cinq étant des chapitres de résultats, et le dernier chapitre étant consacré à la conclusion générale et aux perspectives de mon travail.

Le premier chapitre est donc consacré à une étude bibliographique portant tout d'abord sur les DKP et leur biosynthèse. J'ai ainsi abordé successivement l'abondance naturelle de ces molécules, leur diversité structurale, leurs rôles physiologiques, leurs activités biologiques ainsi que les mécanismes biologiques de leur biosynthèse. L'étude bibliographique a ensuite porté sur la formation biologique des liaisons peptidiques. J'ai choisi de scinder cette partie selon l'implication ou non des aa-ARNt dans la catalyse. Ainsi, dans un premier temps, j'ai décrit les systèmes biologiques impliquant des aa-ARNt et j'ai donc mentionné tout naturellement le système ribosomal de synthèse des protéines avec une présentation brève des ARNt, des aminoacyl-ARNt synthétases et du ribosome. Puis, je me suis attachée à recenser et à décrire les familles d'enzymes qui utilisent des aa-ARNt comme substrats pour former

des liaisons peptidiques. Ces familles d'enzymes sont les aminoacyl-transférases de type Fem qui sont impliquées dans la biosynthèse du peptidoglycane de la paroi des bactéries, les aminoacyl-ARNt protéine transférases qui modifient l'extrémité N-terminale des protéines pour les adresser vers une dégradation protéolytique et, bien sûr, les CDPS qui produisent des cyclodipeptides. Dans un deuxième temps, j'ai décrit les enzymes qui utilisent des acides aminés qui ne sont plus activés sous la forme d'aa-ARNt mais qui sont activés par l'ATP. Ces enzymes sont généralement des ligases formant une seule liaison peptidique, à l'exception des peptides synthétases non ribosomales (NRPS) qui sont à l'origine de la formation de l'essentiel des peptides non ribosomaux. J'ai choisi de présenter plus en détail les connaissances sur les NRPS car ces enzymes sont les seules autres enzymes connues, en plus des CDPS, à synthétiser des cyclodipeptides.

Les chapitres deux à cinq présentent les résultats obtenus pendant mon travail de thèse. Le chapitre 2 est consacré à la caractérisation structurale et mécanistique de la première CDPS identifiée, puis les trois chapitres suivants sont chacun consacrés aux résultats obtenus avec trois autres CDPS, chacune de ces enzymes ayant des caractéristiques particulières et adéquates pour approfondir l'étude de la famille des CDPS.

Le chapitre deux est donc consacré à la première CDPS identifiée, à savoir AlbC de *Streptomyces noursei*. Ce chapitre se compose d'une brève introduction, d'un article (Sauguet et al. 2011) et d'une partie présentant des données additionnelles par rapport à l'article. AlbC utilise essentiellement du Phe-tRNA<sup>Phe</sup> et du Leu-tRNA<sup>Leu</sup> comme substrats pour former majoritairement du cyclo(L-Phe-L-Leu) (cFL) qui est le précurseur de l'albonoursine, cyclo( $\alpha,\beta$ -déshydroPhe- $\alpha,\beta$ -déshydroLeu), une DKP présentant des activités antibiotiques. L'article présente la structure cristallographique de AlbC qui a été déterminée par Ludovic Sauguet, le thésard qui m'a précédée au laboratoire. La structure consiste en un monomère contenant un domaine de type « Rossmann fold ». Ce domaine est très similaire au domaine catalytique des aminoacyl-ARNt synthétases de classe Ic, à savoir les tyrosinyl- et tryptophanyl-ARNt synthétases. AlbC contient une poche profonde et accessible à la surface qui est très conservée chez l'ensemble des CDPS et qui est similaire à la poche qui lie l'acide aminé chez les aminoacyl-ARNt synthétases de classe Ic. AlbC contient également un « patch » de résidus basiques couvrant une surface de 1040 Å<sup>2</sup> que l'on retrouve également chez les autres CDPS mais qui est absent chez les aminoacyl-ARNt synthétases. Par ailleurs, l'article décrit les études de biochimie et de biologie moléculaire entreprises pour identifier et comprendre le rôle des résidus de AlbC dans l'interaction avec les substrats et la catalyse. J'ai ainsi construit un grand nombre de variants de AlbC et analysé leurs activités par un test *in vivo*. Ce

test consiste à exprimer les variants chez *Escherichia coli* et à analyser par LC-MS/MS les surnageants de culture afin d'identifier et de quantifier les cyclodipeptides synthétisés puis secrétés. Cette étude a permis de montrer que la poche de AlbC accommode la partie aminoacyle d'un aa-ARNt d'une manière similaire à celle utilisée par la tyrosyl-ARNt synthétase pour accomoder son substrat tyrosyle. Elle a également indiqué que le « patch » de résidus basiques est important pour l'activité de AlbC, très certainement de par son interaction avec la partie ARNt du substrat. L'article décrit également les avancées obtenues quant au mécanisme catalytique emprunté par AlbC. En effet, nous avons pu montrer que la partie phénylalanyle d'un Phe-ARNt<sup>Phe</sup> se positionne dans la poche de AlbC puis est transférée sur une serine catalytique située à l'entrée de la poche. AlbC utilise donc un mécanisme séquentiel de type ping-pong qui transite par un intermédiaire acyl-enzyme.

Le chapitre 2 se termine par des informations additionnelles consistant essentiellement en une analyse comparative de la structure de AlbC et des structures de deux autres CDPS, Rv2275 de *Mycobacterium tuberculosis* et YvmC de *Bacillus licheniformis*, qui ont été déterminées par deux autres laboratoires (Vetting et al. 2010 ; Bonnefond et al. 2011).

Le chapitre 3 relate les travaux réalisés dans le but d'augmenter nos connaissances sur les bases moléculaires de la spécificité des CDPS. Il se

compose d'une brève introduction, d'un manuscrit d'article et d'une partie brève présentant des données additionnelles par rapport au manuscrit d'article.

Les CDPS présentent entre elles une faible identité de séquences (20-25%) et se distinguent par des profils différents de production de cyclodipeptides. Ainsi, AlbC synthétise essentiellement du cFL mais aussi d'autres cyclodipeptides contenant un résidu phénylalanyle et les autres CDPS caractérisées produisent des cyclodipeptides contenant un résidu tyrosyle, leucyle ou tryptophanyle (Gondry, Sauguet et al. 2009 ; Seguin, Moutiez et al. 2011). Une nouvelle CDPS potentielle a été identifiée chez *Nocardiosis dassonvillei*. Cette CDPS potentielle possède un contexte génomique similaire à celui de AlbC et présente une identité de séquence importante avec AlbC (40%). J'ai montré que Ndas\_1148 est une CDPS active et j'ai établi son profil de production de cyclodipeptides. Ndas\_1148 produit des cyclodipeptides qui sont également produits par AlbC mais leur nombre et leur quantité sont bien moindres, ce qui est cohérent avec le fait que Ndas\_1148 est moins exprimée sous forme soluble que AlbC. De manière tout à fait intéressante, les proportions relatives des cyclodipeptides produits par les deux CDPS ne sont pas identiques. Ainsi, Ndas-1148 synthétise très peu de cyclodipeptides contenant un résidu leucyle. De ce fait, le cFL qui est le produit majoritaire de AlbC n'est quasiment pas produit par Ndas\_1148 qui synthétise majoritairement du cFY et du cFF.

Pour tenter d'identifier les déterminants de la spécificité responsables de la différence des profils de production de cyclodipeptides de Ndas\_1148 et AlbC, un modèle de la structure de Ndas\_1148 a été construit en se basant sur la structure de AlbC. La comparaison de la distribution du potentiel électrostatique de surface des deux CDPS a clairement montré que le « patch » de résidus basiques présent chez AlbC est également présent chez Ndas\_1148. Nous avons précédemment montré que la poche de AlbC, qui est constituée de 19 résidus, lie la partie phénylalanyle du Phe-tRNA<sup>Phe</sup> (Sauguet et al. 2011) mais nous ne disposons d'aucune donnée sur la liaison de l'autre substrat, le Leu-tRNA<sup>Leu</sup>, nécessaire à la synthèse de cFL. La comparaison de cette poche avec celle modélisée pour Ndas\_1148 indique que seulement six résidus sont différents. J'ai alors substitué chez AlbC ces six résidus par ceux présents chez Ndas\_1148, et ainsi obtenu une chimère correspondant à AlbC possédant la poche de Ndas\_1148. La chimère a acquis les profils d'expression protéique et de production de cyclodipeptides de Ndas\_1148. Ce résultat montre clairement que la différence des résidus constituant les poches des deux CDPS est responsable de la différence des profils d'expression des cyclodipeptides. Il montre en particulier que des déterminants de la spécificité responsables de l'incorporation d'un résidu leucyle dans des cyclodipeptides sont localisés dans la poche. J'ai alors introduit indépendamment chez AlbC chacun des six résidus de Ndas\_1148 mais aucun des six variants obtenus n'a acquis un profil similaire à celui de Ndas\_1148. C'est donc l'effet cumulatif des



six substitutions qui est responsable du changement de spécificité. L'ensemble de ce travail, en plus de montrer l'importance de la poche pour la reconnaissance des parties aminoacyles des deux substrats, indique que la reprogrammation de la spécificité d'une CDPS peut être réalisée en échangeant la poche catalytique de cette CDPS par celle d'une autre CDPS.

Le chapitre 3 se termine par des informations additionnelles concernant le travail planifié pour identifier les métabolites finaux produits par *N. dassonvillei* et ainsi caractériser les modifications introduites sur les cyclodipeptides synthétisés par Ndas\_1148.

Le chapitre 4 décrit les travaux préliminaires réalisés sur une autre CDPS, AlbC-IMI de *Streptomyces sp.* IMI 351155, qui présente l'avantage d'utiliser comme substrats du Phe-ARNt<sup>Phe</sup> et du Leu-ARNt<sup>Leu</sup> pour synthétiser quasiment exclusivement du cFL. Cette caractéristique la distingue de AlbC de *S. noursei* qui synthétise majoritairement du cFL mais aussi des quantités importantes de cFF et de cLL. AlbC-IMI est donc un bon modèle pour étudier l'ordre d'interaction des deux substrats et pour déterminer les paramètres cinétiques de chacun de ces deux substrats. Ce chapitre se compose d'une brève introduction qui indique la démarche expérimentale qui a conduit à l'isolement de AlbC-IMI par l'un de nos collaborateurs, d'une partie décrivant

les matériels et les méthodes utilisés, d'une partie mentionnant les résultats et d'une discussion de ces résultats.

Il s'avère que l'obtention de AlbC-IMI entière et pure pour mener à bien nos travaux a nécessité des mises au point expérimentales. En effet, l'utilisation des protocoles classiquement utilisés pour les autres CDPS avait conduit à l'obtention d'une quantité très faible de protéine, qui plus est en partie protéolysée. Le changement de la souche d'expression et l'ajout d'inhibiteurs de protéases adéquats a permis d'obtenir des quantités suffisantes d'une protéine intacte pour mener des expériences *in vitro*.

Il a été précédemment établi que les CDPS utilisent un mécanisme catalytique séquentiel de type ping-pong ; ainsi, un premier aa-ARNt se lie à l'enzyme, la partie aminoacyle de ce substrat est transférée sur une serine située à l'entrée de la poche catalytique pour former une acyl-enzyme puis le deuxième aa-ARNt vient interagir (Sauguet et al. 2011 ; Vetting et al. 2010 ; Bonnefond et al. 2011). J'ai pu détecter la formation d'une phénylalaninyl-enzyme quand AlbC-IMI purifiée est mise en présence de Phe-ARNt<sup>Phe</sup> alors que je n'ai pas pu observer de leucyl-enzyme quand cette même CDPS est incubée avec du Leu-ARNt<sup>Leu</sup>. Ce résultat montre clairement que le Phe-ARNt<sup>Phe</sup> est le premier substrat à interagir avec la CDPS. Par ailleurs, j'ai mis au point les conditions de mesure de l'activité enzymatique *in vitro*, ce qui va permettre d'établir les paramètres cinétiques de chacun des deux substrats.

Ces résultats obtenus avec AlbC-IMI confirment que cette enzyme est un modèle particulièrement intéressant pour l'étude des CDPS et sa caractérisation sera poursuivie au laboratoire.

Le chapitre 5 montre que les CDPS ne sont pas retrouvées uniquement chez des bactéries. En effet, nous avons pu montrer qu'une CDPS potentielle identifiée chez un eucaryote est une CDPS active. Ce chapitre se compose d'une brève introduction, d'un d'article et d'une partie brève présentant des données additionnelles par rapport à l'article.

Ma participation au travail décrit dans l'article (Seguin, Moutiez et al. 2011) a consisté à faire les analyses en LC-MS/MS qui ont permis de mettre en évidence que l'enzyme *Nvec*-CDPS2 identifiée chez l'anémone de mer *Nematostella vectensis* est une CDPS active qui synthétise différents cyclodipeptides. *Nvec*-CDPS2 est la première enzyme impliquée dans la synthèse peptidique non ribosomale à être identifiée chez un animal. Ce résultat soulève de nombreuses questions quant aux rôles physiologiques des cyclodipeptides produits chez des eucaryotes.

Par ailleurs, nous avons pu montrer que, à l'instar des membres bactériens de la famille des CDPS, *Nvec*-CDPS2 possède une structure avec un domaine de type « Rossmann fold » qui ressemble au domaine catalytique des aminoacyl-ARNt synthétases de classe Ic. Nous avons également établi que

*Nvec*-CDPS2 utilise un mécanisme séquentiel de type ping-pong transitant par une acyl-enzyme intermédiaire. Par contre, *Nvec*-CDPS2 ne possède pas la même spécificité de substrats que les autres CDPS déjà caractérisée car elle produit des cyclodipeptides contenant préférentiellement un résidu tryptophanyle. Cela laisse supposer que d'autres CDPS avec d'autres spécificités de substrats restent à découvrir.

En données additionnelles, je mentionne que d'autres CDPS potentielles ont été identifiées chez des eucaryotes (Aravind et al. 2010 ; Belin, Moutiez et al., 2012).

Le chapitre 6 de ce manuscrit correspond à une partie présentant les conclusions et les perspectives de mon travail de thèse.

Je rappelle donc que, après la présentation d'une étude bibliographique sur les DKP et les biocatalyseurs impliqués dans la formation de liaisons peptidiques (Chapitre 1), j'ai tout d'abord décrit ma participation à l'étude tant structurale que mécanistique de la première CDPS identifiée, AlbC de *Streptomyces noursei*, (Chapitre 2). Puis, je mentionne les résultats obtenus avec trois autres CDPS, chacune de ces enzymes ayant des caractéristiques adéquates pour approfondir l'étude de la famille des CDPS. Ainsi, la CDPS Ndas\_1148 de *Nocardioopsis dassonvillei* a permis d'étendre nos connaissances sur les bases moléculaires de la spécificité des CDPS

(Chapitre 3). La CDPS AlbC-IMI de *S. sp.* IMI 351155 est un bon modèle pour analyser l'interaction de chacun des deux substrats nécessaires à la formation d'un cyclodipeptide (Chapitre 4). Enfin, la caractérisation de la CDPS *Nvec*-CDPS2 chez l'animal *Nematostella vectensis* a permis de fournir le premier exemple d'enzyme d'origine animale impliquée dans la synthèse peptidique non ribosomale (Chapitre 5).

A ce jour, des CDPS potentielles ont été identifiées dans environ 60 génomes de bactéries, de champignons ou d'animaux (Aravind et al. 2010; Belin, Moutiez et al. 2012). Ainsi de nombreuses CDPS restent à être caractérisées. De plus, ces gènes de CDPS étant très souvent associés à des gènes codant des enzymes de modification des cyclodipeptides, de nombreuses enzymes de modification restent aussi à être caractérisées. L'ensemble de ces caractérisations ouvrent la voie à de l'ingénierie des voies de biosynthèse de DKP dans le but de produire de nouvelles DKP aux propriétés biologiques et pharmacologiques d'intérêt.

*Gondry M., Sauguet L., Belin P., Thai R., Amouroux R., Tellier C., Tuphile K, Jacquet M., Braud S., Courçon M., Masson C., Dubois S., Lautru S., Lecoq A., Hashimoto S.-i, Genet R. and Pernodet J.-L. (2009). "Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes." Nature Chemical Biology **5**: 414-420.*

*Belin P., Le Du M.-H., Fielding A., Lequin O, Jacquet M., Charbonnier J.-B., Lecoq A., Thai R., Courçon M., Masson C., Dugave C., Genet R., Pernodet J.-L. and Gondry M.*

(2009). "Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*." Proceedings of the National Academy of Sciences of the United States of America **106**: 7426-7431.

Belin P., Moutiez M., Lautru S., Seguin J., Pernodet J.-L. and Gondry M. (2012). "The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways." Natural Product Reports **29**: 961-979.

Sauguet L., Moutiez M., Li Y., Belin P., Seguin J., Le Du M.-H., Thai R., Masson C., Fonvielle M., Pernodet J.-L., Charbonnier J.-B., Gondry M. (2011). "Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis." Nucleic Acids Research **39**: 4475-4489.

Vetting M. W., Hegde S. S. and Blanchard J. S. (2010). "The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase." Nature Chemical Biology **6**: 797-799.

Bonnefond L., Arai T., Sakaguchi Y., Suzuki T., Ishitani R. and Nureki O. (2011). "Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog." Proceedings of the National Academy of Sciences of the United States of America **108**: 3912-3917.

Seguin J., Moutiez M., Li Y., Belin P., Lecoq A., Fonvielle M., Charbonnier J.-B., Pernodet J.-L., Gondry M. (2011). "Nonribosomal peptide synthesis in animals: the cyclodipeptide synthase of *Nematostella*." Chemistry & Biology **18**: 1362-1368.

Aravind L., de Souza R. F. and Iyer L. M. (2010). "Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis." Biology Direct **5**(48): 48-58.