



**HAL**  
open science

# Étude multidisciplinaire des aspects clés de la biosynthèse des polykétides par des polykétide synthases modulaires

Thibault Annaval

## ► To cite this version:

Thibault Annaval. Étude multidisciplinaire des aspects clés de la biosynthèse des polykétides par des polykétide synthases modulaires. Biochimie, Biologie Moléculaire. Université de Lorraine, 2015. Français. NNT : 2015LORR0241 . tel-01754505

**HAL Id: tel-01754505**

**<https://hal.univ-lorraine.fr/tel-01754505v1>**

Submitted on 30 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

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

**Ecole Doctorale BioSE (Biologie-Santé-Environnement)**

**Thèse**

**Présentée et soutenue publiquement pour l'obtention du titre de**

**DOCTEUR DE L'UNIVERSITE DE LORRAINE**

**Mention : « Sciences de la Vie et de la Santé » par**

**Thibault ANNAVAL**

**Etude multidisciplinaire des aspects clés de la biosynthèse  
des polykétides par des polykétide synthases modulaires**

**Le 17 décembre 2015**

**Membres du jury :**

<b>Rapportrices :</b>	<b>Mme Coralie Bompard</b>	<b>Chargée de recherche, CNRS, Lille</b>
	<b>Mme Muriel Gondry</b>	<b>Chercheure CEA, CEA, Saclay</b>
<b>Examineurs :</b>	<b>Mme Kira J Weissman</b>	<b>Professeure, Université de Lorraine, Nancy, directrice de thèse</b>
	<b>M. Christophe Jacob</b>	<b>Maître de conférences, Université de Lorraine, Nancy, co-directeur de thèse</b>
<b>Membre invité:</b>	<b>M. Bertrand Aigle</b>	<b>Professeur, Université de Lorraine, Nancy</b>

---

**Laboratoire IMoPA: Ingénierie Moléculaire et Physiopathologie Articulaire  
UMR 7365 CNRS-UL, Faculté de Médecine, Bâtiment Biopôle,  
9 Avenue de la Forêt de Haye, BP 184, 54505 Vandoeuvre-lès-Nancy**

*A Saccharopolyspora erythraea...*

# *Table des matières*

## Table des matières

Abréviations .....	1
Introduction .....	4
I. Les métabolites secondaires .....	4
II. Les polykétides .....	5
III. Les polykétide synthases (PKS) .....	7
III.1 Découverte et organisation génique .....	7
III.2 Classification et fonctionnement .....	10
III.2.1 Les PKS de type III .....	11
III.2.2 Les PKS de type II .....	12
III.2.3 Les PKS de type I .....	13
III.2.3.1 Les PKS de type I itératives .....	13
III.2.3.2 Les PKS de type I modulaires .....	14
III.2.3.2.1 Le type <i>cis</i> -AT .....	14
III.2.3.2.2 Le type <i>trans</i> -AT .....	16
III.2.3.2.3 Différence évolutive des PKS de type <i>cis</i> et <i>trans</i> -AT .....	18
III.3 Ingénierie des polykétide synthases .....	20
IV. Domaines fonctionnels des PKS de type I .....	22
IV.1 Les domaines acyltransférase .....	22
IV.2 Les domaines <i>acyl carrier protein</i> .....	23
IV.3 Les domaines kétosynthase .....	24
IV.4 Les domaines kétoréductase .....	27
IV.5 Les domaines déshydratase .....	28
IV.6 Les domaines énoyl réductase .....	30
IV.7 Les domaines méthyltransférase .....	31
IV.8 Les domaines thioestérase .....	32
V. Interactions inter-domaines .....	34
V.1 Communications intra-modulaires .....	34
V.2 Etude du stéréocontrôle .....	36
V.2.1 Les domaines candidats au stéréocontrôle .....	37
V.2.1.1 Les domaines acyltransférase .....	37
V.2.1.2 Les domaines kétosynthase .....	38

V.2.1.3	Les domaines kétéroréductase .....	38
V.2.2	Les kétéroréductases.....	39
V.2.2.1	Activité réductase des kétéroréductases.....	39
V.2.2.2	Activité épimérase des kétéroréductases.....	43
V.2.3	Stabilisation de la stéréochimie du polykétide .....	46
VI.	Structure et dynamique des modules.....	47
VII.	Les domaines de docking .....	53
VII.1	Les domaines de docking de classe 1a.....	55
VII.2	Les domaines de docking de classe 1b.....	57
VII.3	Les domaines de docking de classe 2.....	59
VII.	Le processus de maturation post-PKS .....	60
	Objectifs .....	63
	Matériels et méthodes.....	65
I.	Matériels biologiques .....	65
I.1	Souches bactériennes.....	65
I.2	Vecteurs plasmidiques.....	65
I.3	Cosmide.....	66
I.4	Milieux de culture .....	66
II.	Clonage, production et purification des protéines recombinantes .....	68
II.1	Extraction de l'ADN génomique de différentes espèces de <i>Streptomyces</i> .....	68
II.2	Clonage des ORF codant les protéines étudiées.....	68
II.2.1	Amplification, purification et ligation.....	68
II.2.2	Transformation bactérienne et sélection des vecteurs recombinants .....	70
II.2.3	L'extraction d'ADN plasmidique.....	71
II.2.4	Sous-clonage des ORF codant les protéines étudiées .....	71
II.3	Expression des protéines .....	72
II.4	Purification des protéines .....	73
II.4.1	Lyse cellulaire .....	73
II.4.2	Chromatographie d'affinité Nickel-Sépharose.....	73
II.4.3	Protéolyse du tag histidine .....	74
II.4.4	Chromatographie d'affinité Nickel-Sépharose inverse .....	74
II.4.5	Filtration sur gel ou chromatographie d'exclusion de taille.....	75
II.4.6	Dosage des protéines .....	75

III. Vers une structure des domaines KR .....	76
III.1 Analyse de l'homogénéité de la protéine .....	76
III.2 Cristallisation .....	76
III.3 Congélation des cristaux obtenus .....	78
III.4 Enregistrement et traitement des données .....	78
III.5 Obtention de dérivés d'atomes lourds de cristaux de protéine.....	79
IV. Etude des interactions entre le substrat CoA-dikétide et les domaines ACP, KS-AT et KR du premier module de la PKS pikromycine.....	80
IV.1 Synthèse du CoA-dikétide.....	80
IV.2 Modification de l'ACP .....	81
IV.3 Acquisition des spectres RMN.....	82
Résultats .....	85
Partie 1 : Caractérisation d'une nouvelle classe de domaine de docking de PKS trans-AT.....	85
Partie 2 : Conversion d'un module non-épimérisant en un module épimérisant par échange de domaine KR.....	130
Partie 3 : Essai de résolution de la structure d'une KR épimérisante .....	171
I. Clonage et purification des KR .....	171
II. Cristallisation des KR.....	174
II.1 Etapes de criblage et d'amélioration des premiers cristaux .....	174
II.2 Utilisation d'additifs.....	175
II.3 Micro-ensemencement .....	176
III. Enregistrement et traitement des données diffraction .....	177
IV. Phasage .....	177
IV.1 Remplacement moléculaire .....	177
IV.2 Obtention de dérivés lourds.....	178
Partie 4 : Etude de la stabilisation de la stéréochimie du C $\alpha$ .....	180
I. Clonage et purification de PIKS ACP <sub>1</sub> , KS <sub>1</sub> -AT <sub>1</sub> et KR <sub>1</sub> .....	180
I.1 Clonage et purification du domaine ACP <sub>1</sub> .....	180
I.2 Clonage et purification du didomaine KS <sub>1</sub> -AT <sub>1</sub> .....	181
I.3 Clonage et purification du domaine KR <sub>1</sub> .....	182
II. Attribution de la chaîne principale de l'ACP <sub>1</sub> .....	183



III. Formation de la forme <i>holo</i> de l'ACP <sub>1</sub> et analyse de la modification .....	184
III.1 Réaction de modification .....	184
III.2 Confirmation de la modification par spectrométrie de masse .....	184
III.3 Analyse de la position de modification de l'ACP <sub>1</sub> .....	186
III.4 Analyse des modifications par RMN en utilisant la Sfp .....	188
IV. Acylation de l'ACP <sub>1</sub> par le CoA-dikétide et analyse de la modification ...	190
IV.1 Synthèse des substrats .....	191
IV.1.1 Purification des enzymes de synthèse du substrat.....	191
IV.1.2 Synthèse du substrat .....	191
IV.1.3 Analyse par spectrométrie de masse de l'obtention des CoA-dikétides .....	192
IV.2 Modification de l'ACP <sub>1</sub> par le CoA-dikétide et analyse par spectrométrie de masse .....	193
IV.2.1 Modification de l'ACP <sub>1</sub> .....	193
IV.2.2 Vérification de la modification par spectrométrie de masse .....	194
Discussion/perspectives .....	199
Références bibliographiques .....	209

# *Abréviations*

## **Abréviations**

6-dEB : 6-déoxyérythronolide B

ACP : *acyl carrier protein*

ADN : acide désoxyribonucléique

Amph : amphotéricine

AT : acyltransférase

ATP : adénosine triphosphate

$\beta$ -ME :  $\beta$ -mercaptoéthanol

BSA : albumine de sérum bovin

CD : dichroïsme circulaire

CoA : coenzyme-A

Con : concanamycine

Cter : extrémité C-terminale

Da : dalton

DEBS : 6-déoxyérythronolide B synthase

DD : domaine de docking

DH : déshydratase

DLS : diffusion dynamique de la lumière

DO : densité optique

DO<sub>3</sub>A : acide 1,4,7,10-tetraazacyclododecane-1,4,7-triacétique

DOTA : acide 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetra-acétique

DPA<sub>3</sub> : tris(dipicolinate)

DPCK : 3'-déphosphocoenzymeA kinase

DTT : dithiothréitol

Elai : elaiophiline

ER : énoyl réductase

FAS : acide gras synthase

GC-MS : chromatographie en phase gazeuse couplée à de la spectrométrie de masse

HC : hétérocyclases

HPDO<sub>3</sub>A : acide 10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacétique

HEPES : acide 4-(2-hydroxyéthyl)-1- pipérazine éthane sulfonique

IDP : protéine intrinsèquement désordonnée

KR : kétoréductase

KS : kétosynthase

Lank : lankamycine

LC-MS : chromatographie liquide couplée à de la spectrométrie de masse

MAD : dispersion anormale à multiple longueur d'onde

MAT : malonyl-acétyl transférase

MCAT : malonyl-CoA transférase

MES : acide 2-(N-morpholino) éthane sulfonique

MLS : mycolactone synthase

MT : méthyltransférase

NADPH : Nicotinamide adénine dinucléotide phosphate

NRPS : synthase de peptides non ribosomiques

Nter : extrémité N-terminale

ORF : cadre ouvert de lecture

PanK : pantéthéine kinase

PCP : *peptidyl carrier protein*

PCR : réaction de polymérisation en chaîne

PDB : banque de donnée des protéines

PdI : polydispersité

Pik : pikromycine

PKS : polykétide synthase

PPAT : phosphopantéthéine adénylyltransférase

PPTase : phosphopantéthéinyl transférase

RMN : résonance magnétique nucléaire

SAD : dispersion anormale à une seule longueur d'onde

SAM : *S*-adénosylméthionine

SAXS : diffusion des rayons X aux petits angles

SDS-PAGE : électrophorèse en gel de polyacrylamide contenant du dodécylsulfate de sodium

Sfp : Surfactine phosphopantéthéinyl-transférase

Svp : *Streptomyces verticillus* PPTase

TCEP : tris(2-carboxyethyl)phosphine

TE : thioestérase

TEV : protéase du virus de la gravure du tabac

TFE : Trifluoroéthanol

T<sub>m</sub> : température d'hybridation

Vir : virginiamycine

# *Introduction*

## Introduction

Bien que l'usage eût été d'utiliser les termes français (comme par exemple polycétide synthase) se rapportant à cette thématique de recherche, nous avons décidé de franciser les termes anglais (polykétide synthase) dans ce manuscrit.

### I. Les métabolites secondaires

Les métabolites secondaires sont retrouvés chez les bactéries, les champignons, les plantes ou encore les animaux. Comme leur nom l'indique, ces molécules naturelles sont des composés non impliqués dans le métabolisme primaire ; c'est-à-dire n'intervenant pas dans la pérennité de l'espèce qui les synthétise que ce soit au niveau de sa reproduction ou bien de son développement. Elles vont permettre une certaine adaptabilité de cette dernière à son environnement. Cependant les métabolites primaires et secondaires ne sont pas complètement distincts. En effet, les voies de biosynthèse des métabolites secondaires utilisent des métabolites et des intermédiaires provenant du métabolisme primaire (Figure 1).

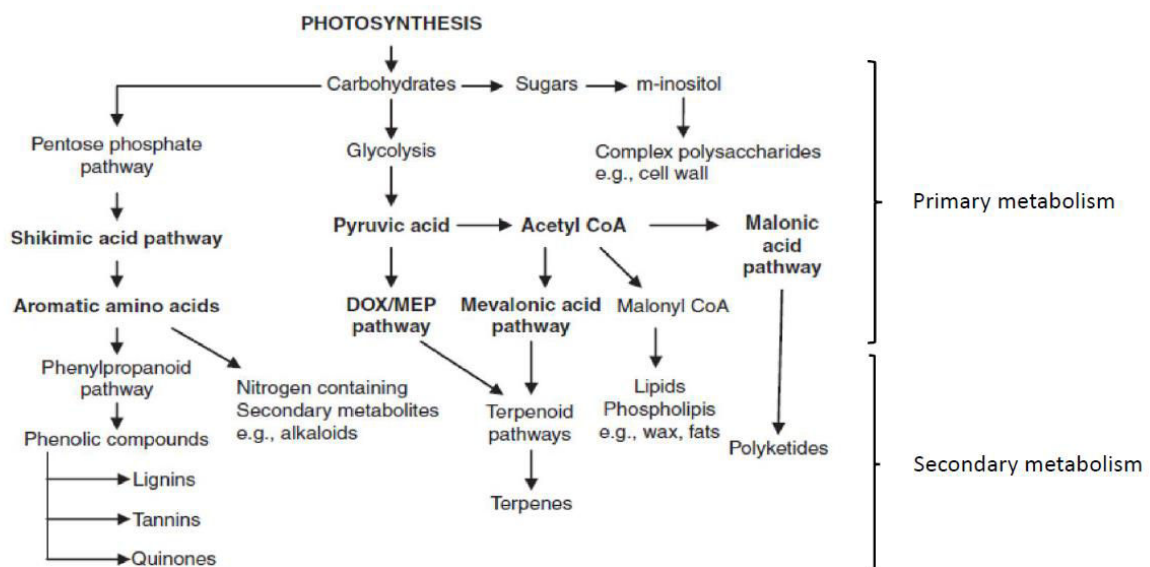


Figure 1 : Voies de biosynthèse des métabolites secondaires découlant des glucides (métabolisme primaire) chez la plante<sup>1</sup>.

Ces métabolites secondaires peuvent se diviser en six catégories : les terpènes/stéroïdes, les glucides spéciaux, les phénylpropanoïdes, les alcaloïdes, les acides aminés et peptides non-protéinogènes et enfin les polykétides<sup>2</sup>.

A ce jour, environ un million de produits naturels a été découvert<sup>3</sup>. Parmi cette quantité de métabolites, 25% présentent une activité biologique, qu'elle soit bénéfique ou toxique pour l'Homme. Cette dernière proportion de molécules provient majoritairement des plantes et des micro-organismes<sup>3</sup>. L'ensemble des métabolites secondaires qui exhibent des activités biologiques bénéfiques continue de représenter une source importante de nouveaux composés, même en face des avancées synthétiques. En effet, entre 1985 et 2005, 50% des médicaments nouvellement introduits sur le marché étaient des produits naturels<sup>3</sup>. Ainsi, de leur découverte jusqu'à leur utilisation thérapeutique croissante, les métabolites secondaires ont permis à l'espérance de vie d'être doublée au cours du 20<sup>ème</sup> siècle<sup>3</sup>.

## **II. Les polykétides**

Les polykétides constituent une grande classe de métabolites secondaires issus des bactéries, plantes et champignons. Ces composés sont caractérisés par leurs activités biologiques (Figure 2), que ce soit antibiotique (érythromycine A), immunosuppresseur (rapamycine), anticancéreux (épothilone), antifongique (amphotéricine) ou encore anticholestérolémique (lovastatine). Les produits pharmaceutiques dérivés de polykétides rapportent énormément à l'industrie pharmaceutique. Par exemple, six polykétides découverts récemment et utilisés en thérapeutique (la lovastatine, la pravastatine, la simvastatine, la clarithromycine, l'azithromycine et le sirolimus) rapportent chaque année 15 milliards de dollars (Natural product pharmaceuticals - the third generation, McCullagh, 2008). Les polykétides jouent donc un rôle important dans le domaine médical et présentent un grand intérêt économique.

En plus de leur grande diversité de fonction, les polykétides présentent une riche variété de structures (Figure 2). En regard des molécules de synthèse utilisées en médecine, ils possèdent une complexité structurale bien plus élevée<sup>4</sup> (Figure 3). En règle générale, les polykétides sont des molécules rigides caractérisées par une densité de groupements fonctionnels élevée ainsi que par une stéréochimie bien définie.



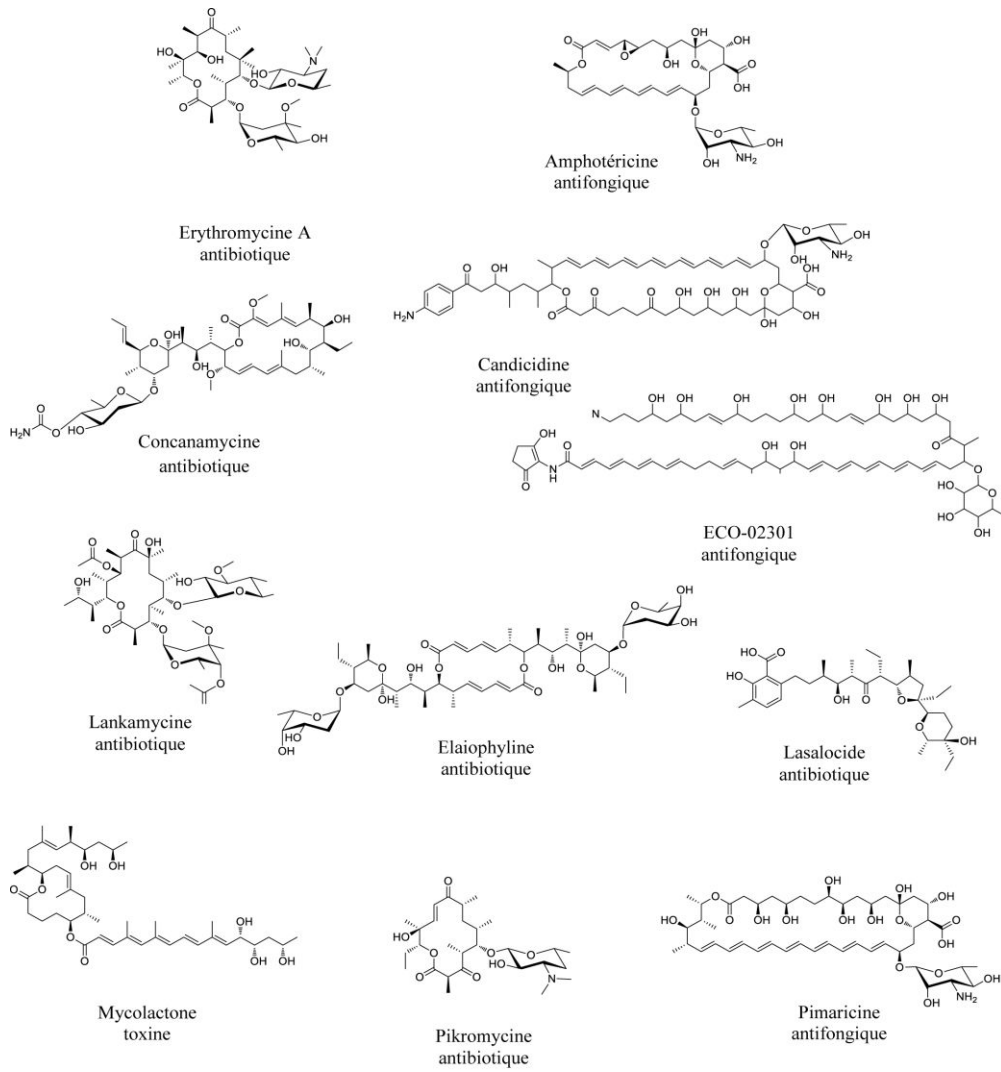


Figure 2 : Exemples de la complexité et de la diversité des polykétides et de leur fonction associée.

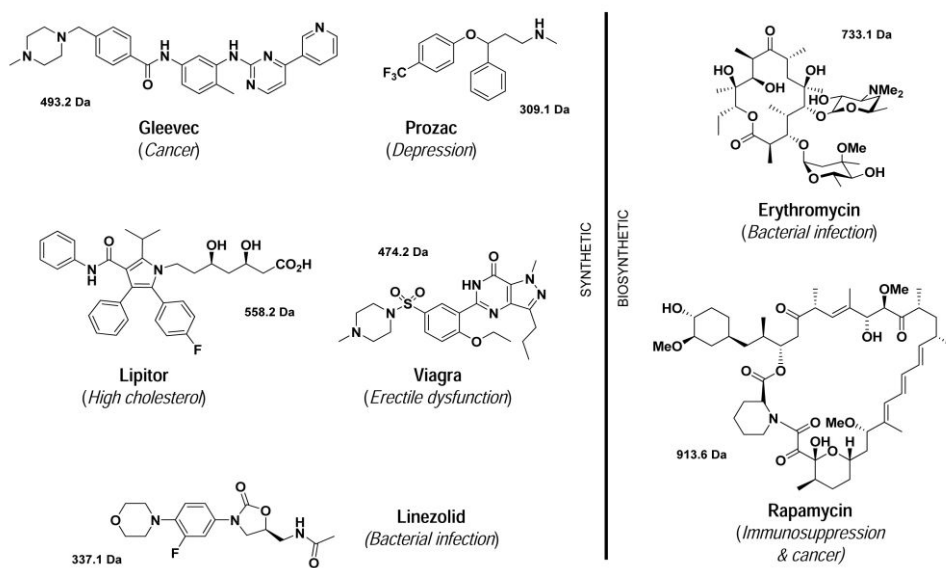


Figure 3 : Comparaison des molécules issues de synthèse et de biosynthèse utilisées en médecine<sup>4</sup> (adapté par KJ Weissman).

La découverte de nouvelles structures de polykétides s'avère donc être une des stratégies intéressantes dans le but d'obtenir de nouvelles molécules à potentiel médical. Pour ce faire, trois approches sont possibles : la découverte de nouveaux polykétides dans la biosphère, la synthèse/hémi-synthèse chimique (la modification des structures existantes par la chimie synthétique) ou l'ingénierie génétique des voies de biosynthèse, une approche désignée sous le terme de « biologie synthétique »<sup>5</sup>. Cette dernière approche semble être prometteuse pour aboutir à la découverte de nouveaux polykétides, même si les deux premières restent les plus efficaces à présent. Cependant, elle nécessite de comprendre en détail les chemins biosynthétiques afin d'envisager une ingénierie efficace.

### III. Les polykétide synthases (PKS)

Les polykétides sont biosynthétisés par un ensemble de protéines. Les polykétide synthases, sont responsables de la biosynthèse du squelette du polykétide. Ensuite des enzymes dites « post-PKS », avec des activités très variées, réalisent la modification du squelette de manière à former un polykétide actif.

Les PKS sont des enzymes qui réalisent les différentes étapes d'assemblage des polykétides *via* la condensation de simples blocs de construction, des unités acyl-CoA. Elles sont composées d'un ou plusieurs domaines, qui permettent de sélectionner les blocs de construction qui serviront à la synthèse des polykétides, puis de les incorporer dans la chaîne en croissance et de les modifier. La manière chimique dont sont synthétisés les polykétides est fortement homologue à la synthèse des acides gras, la lipogénèse, par les acide gras synthases (FASs : Fatty Acid Synthases). Cette homologie de mécanisme est corrélée à une homologie au niveau de la séquence primaire de ces deux types de synthases qui démontre qu'elles possèdent une origine commune en terme d'évolution.

#### III.1 Découverte et organisation génique

Le développement du génie génétique et de ses outils a permis de caractériser l'organisation génique des PKS. Les travaux réalisés sur l'ADN génomique de *Streptomyces coelicolor*, souche responsable de la synthèse de l'actinorhodine, ont permis de mettre en évidence l'organisation en cluster des gènes codant les PKS. Tout d'abord, des mutations au niveau de l'ADN génomique ont permis de localiser la partie

génique responsable de la biosynthèse de l'actinorhodine par observation de la perte de la couleur bleue caractéristique du polykétide au niveau des colonies obtenues<sup>6</sup>. Ensuite, des mutations ont été réalisées à l'intérieur de la partie génique identifiée précédemment et six phénotypes directement liés à la biosynthèse du polykétide ont été mis en évidence. Chacun de ces phénotypes résulte de mutations localisées au sein d'un même locus, suggérant la présence d'au moins six gènes intervenant au niveau de la synthèse de l'actinorhodine et regroupés en cluster<sup>7</sup>.

Par la suite, le cluster a été séquencé entièrement et a permis l'obtention de la structure primaire de la PKS<sup>8</sup> (Figure 4). Sa composition en domaine a été déterminée par comparaison avec des protéines homologues et notamment les FAS, au niveau de trois protéines discrètes qui collaborent pour synthétiser le polykétide.

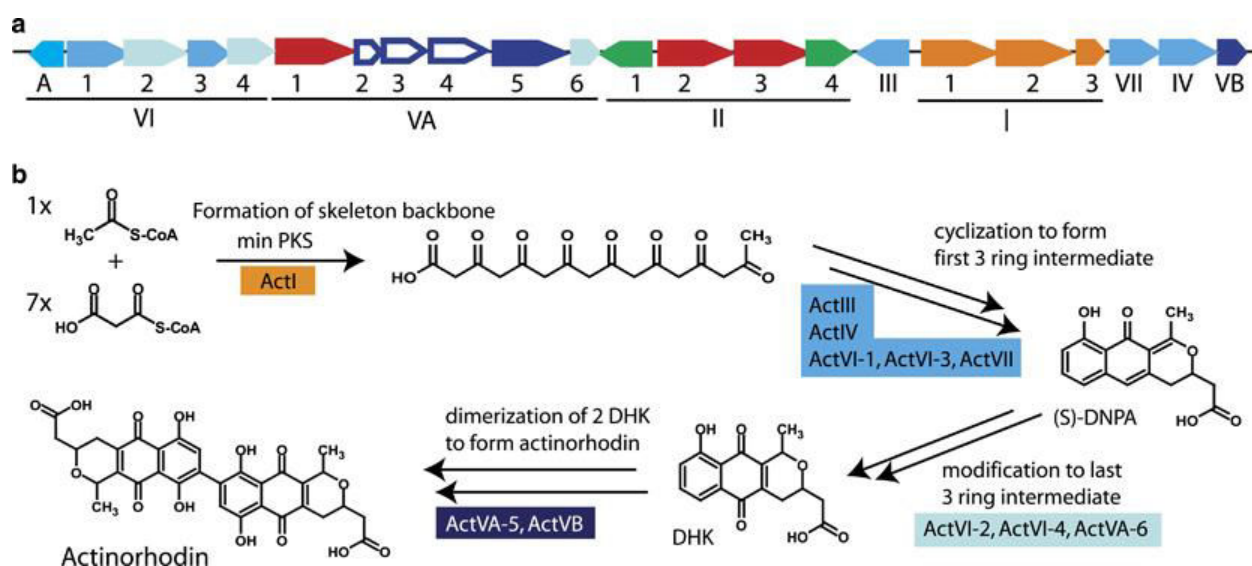


Figure 4 : a : Cluster de gène de la PKS biosynthétisant l'actinorhodine. Les gènes codant les trois protéines qui constituent la PKS sont en orange, ceux de régulation en vert, ceux de résistance à l'actinorhodine en rouge et ceux qui codent les protéines de modification en bleu. b : Biosynthèse de l'actinorhodine. Les protéines codées par les gènes du cluster sont indiquées pour chaque étape de la biosynthèse<sup>8</sup>.

Suite à l'identification de ce cluster de gènes, des études ont portées sur l'identification du cluster de gènes de la PKS biosynthétisant l'antibiotique érythromycine A, qui deviendra par la suite un modèle d'étude. Pour se faire, deux types d'approches ont été utilisées<sup>9,10</sup>. Toutes deux sont basées sur l'hypothèse selon laquelle le cluster de gènes responsable de la biosynthèse du polykétide se situe à proximité du gène responsable de son auto-résistance. L'érythromycine agit en tant qu'antibiotique en se liant à la sous-unité 50S du ribosome, inhibant ainsi la croissance

bactérienne. Pour se préserver de son polykétide, le cluster codant l'érythromycine contient le gène *ermE* codant une ARNr méthylase, qui catalyse une méthylation, chez *Saccharopolyspora erythraea*<sup>11</sup>.

Pour identifier le cluster, une première étude s'est basée sur les travaux concernant l'actinorhodine. Une région au sein du génome de *S. erythraea* a été mise en exergue car elle provoquait l'inhibition de la biosynthèse du polykétide une fois mutée. Les mutations ont été effectuées à proximité de la région qui code la résistance à l'érythromycine A. Cette zone du génome bactérien a été séquencée de part et d'autre et a permis d'identifier les gènes de la PKS<sup>9</sup>. La deuxième approche, également basée sur la région de résistance à l'érythromycine A, a consisté à découvrir le cluster de gènes en parcourant le chromosome à proximité du gène *ermE* et en cherchant des séquences possédant une forte similarité avec d'autres PKS (de type II notamment) ou bien des FAS<sup>10</sup>.

Le cluster de gènes responsable de la synthèse de l'érythromycine a donc été caractérisé. Il s'étend sur 56 kb. Il est composé de trois ORF *eryAI*, *eryAII* et *eryAIII* codant les trois sous-unités qui forment la PKS, elles-mêmes composées de plusieurs domaines. Ce cluster comprend évidemment le gène *ermE*, ainsi que les gènes codant des enzymes intervenant lors des modifications post-PKS du polykétide qu'il s'agisse de la méthylase, des hydroxylases, des glycosyltransférases et des enzymes intervenant dans la biosynthèse de ses glucides transférés (Figure 5).

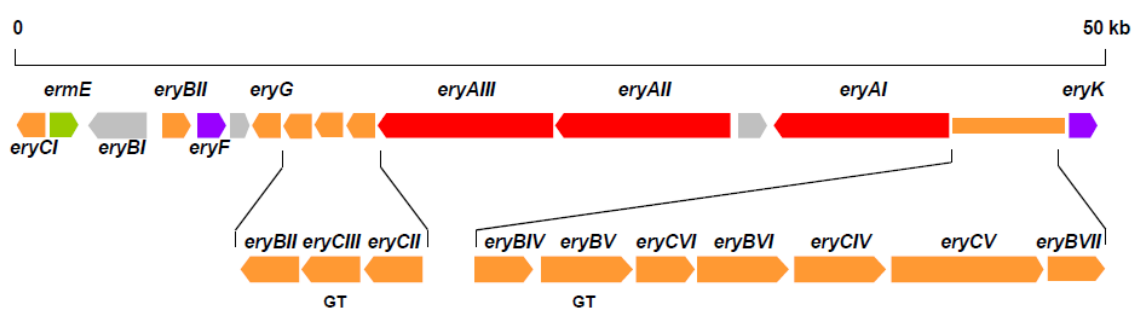


Figure 5 : Représentation du cluster de gènes responsable de la synthèse de l'érythromycine A chez *S. erythraea*. En rouge sont indiqués les gènes de la PKS, en orange les gènes des enzymes de la synthèse des glucides, des glycosyltransférases et de la méthylase, en violet les gènes des hydroxylases et en vert, le gène conférant la résistance (Weissman KJ).

### III.2 Classification et fonctionnement

L'analyse du cluster de la PKS biosynthétisant l'érythromycine A (DEBS) a tout de suite mis en avant que cette classe de polykétides est synthétisée par un type d'enzyme différent de celui biosynthétisant l'actinorhodine. En effet, DEBS est composée de domaines fonctionnels regroupés au sein de gigantesques multienzymes alors que la PKS de l'actinorhodine est composée de trois protéines discrètes. Plusieurs types de PKS sont alors définis : le type I dont DEBS fait partie, le type II avec la PKS de l'actinorhodine et enfin le type III qui est composé d'une seule protéine qui réalise toute les tâches d'assemblage. Ce classement tient compte de la nature des produits obtenus : les types II et III conduisent à des polykétides aromatiques tandis que le type I permet d'obtenir des polykétides réduits ou complexes que ce soit des macrocycles, des polyéthers ou des polykétides linéaires<sup>12</sup>. Enfin le mode opératoire, de type modulaire ou itératif, définit également les catégories de PKS.

Les PKS de type I et II présentent des caractéristiques similaires avec les enzymes de la biosynthèse des acides gras, contrairement au type III. Ainsi, avant de commencer la description de chaque type de PKS et son fonctionnement, il est approprié de mettre encore un peu plus en avant l'origine commune des PKS et des FAS, en décrivant la synthèse des acides gras.

La lipogénèse est un ensemble de réactions intervenant au sein du métabolisme primaire. Il s'agit d'un processus permettant la formation d'acides gras *via* la condensation répétée de deux composés réalisée grâce aux acides gras synthases (FASs). Tous les types de FASs sont composés de plusieurs domaines communs : malonyl-acétyl transférase (MAT), *acyl carrier protein* (ACP), kétosynthase (KS), kétoréductase (KR), déshydratase (DH), énoyl réductase (ER) et thioestérase (TE). Au sein des bactéries, ces domaines sont séparés (FAS de type II) tandis que chez les mammifères ils sont liés coalement et forment ainsi des protéines multifonctionnelles (FAS de type I). La synthèse des acides gras débute par le chargement d'un acétyl-CoA sur le domaine ACP *via* le domaine MAT. L'acétyl est ensuite lié au domaine KS pour être condensé avec une molécule d'extension, le malonate, qui est également choisit par le domaine MAT et lié à l'ACP. Puis il y a modification de l'unité d'extension nouvellement ajoutée par l'action séquentielle des domaines KR, DH et ER. Le domaine KR réalise la réduction de la fonction cétone en  $\beta$  pour donner un hydroxyle. Le domaine DH réalise quant à lui l'élimination de l'hydroxyle *via* la formation d'une double liaison. Enfin le domaine ER réduit la double

liaison formée par le domaine DH pour donner une liaison simple. La chaîne en cours de croissance est agrandie de manière itérative, par ajout de nouvelles molécules de malonate, jusqu'à obtention de l'acide gras de longueur approprié qui sera alors libéré par le domaine TE (Figure 6).

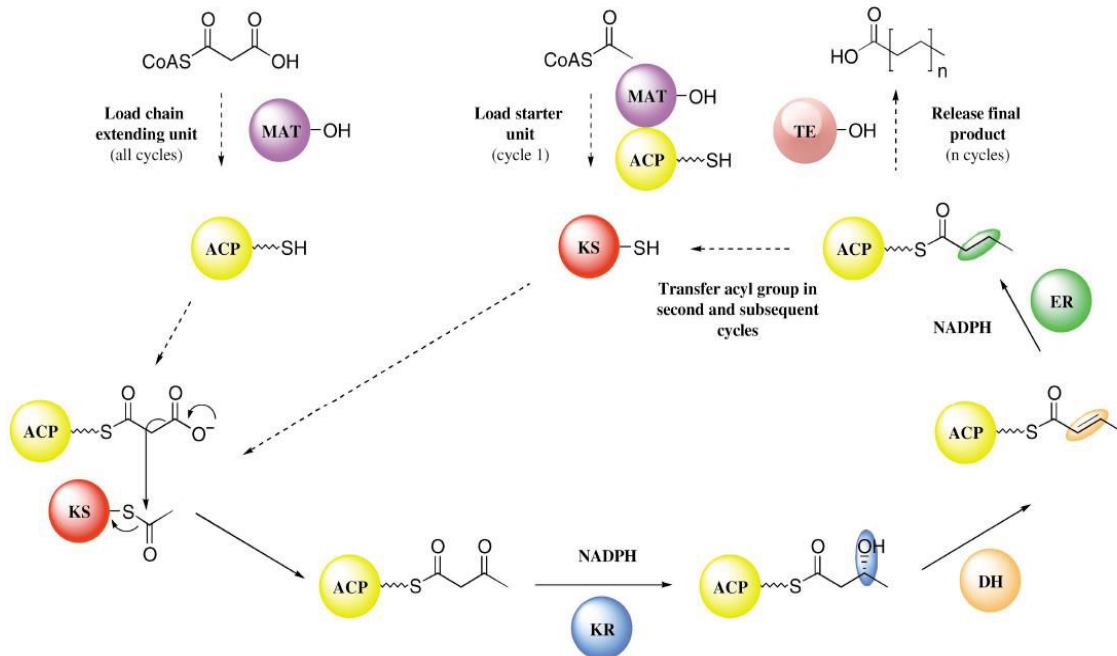
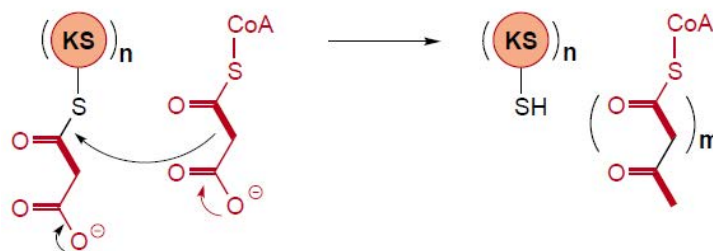


Figure 6 : Représentation de la synthèse des acides gras<sup>13</sup>. MAT : malonyl-acétyl transférase, ACP : *acyl carrier protein*, KS: kétosynthase, KR : kétoréductase, DH : déshydratase, ER : énoyl réductase, TE: thioestérase.

### III.2.1 Les PKS de type III

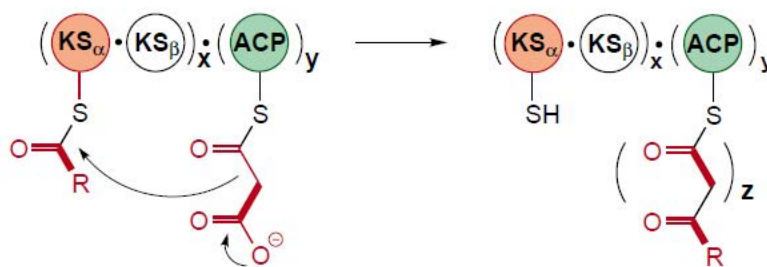
Ce type de PKS, également appelée chalcone et stilbène synthases, est ubiquitaire chez les plantes ; elle intervient au sein de la première étape de la biosynthèse des flavonoïdes<sup>14</sup>. Néanmoins, elle est également présente chez les bactéries et champignons. Ce type d'enzyme montre une architecture simple, c'est-à-dire composée d'une kétosynthase homodimérique multifonctionnelle, contrairement aux PKS de type I et II qui présentent une organisation multienzymatique. Les PKS de type III sont dites itératives puisqu'elles utilisent ce seul site actif de KS pour réaliser la condensation répétée de plusieurs acyl-CoA afin de former le polykétide. A la fin de ces condensations répétées, il se produit une étape de cyclisation puis une étape de modification du produit formé *via* des enzymes post-PKS. Contrairement aux PKS de type I et II pour lesquelles les unités d'extension qui seront condensées sont fixées au niveau d'une ACP, les unités d'extension au sein des PKS de type III restent chargées sur des molécules de coenzyme A<sup>14</sup> (Figure 7).

Figure 7 : Représentation et mode d'action d'une PKS de type III<sup>15</sup>.

### III.2.2 Les PKS de type II

Ce type de PKS synthétise notamment une des plus grandes classe d'antibiotiques que sont les tétracyclines. A l'instar des PKS de type III, ces enzymes sont également itératives. Cependant, celles-ci présentent une organisation plus complexe, de type multienzymatique<sup>15</sup>. Elles possèdent une forte homologie avec les acides gras synthases bactériennes que ce soit en terme de mécanisme de l'extension de la chaîne mais également en terme de composition en certains domaines.

Un assortiment d'enzymes composant une « PKS minimale » est nécessaire afin de permettre la synthèse des polykétides (Figure 8). A cette PKS minimale peuvent s'ajouter des fonctions enzymatiques supplémentaires telles que des kétéductases, cyclases ou encore aromatasés qui réalisent la maturation de la chaîne en cours de croissance. La synthèse se réalise par la condensation itérative d'unités malonyl-CoA *via* les deux KS : le domaine  $KS_{\beta}$  fournit l'unité de départ, l'acétyl-CoA, obtenu par décarboxylation du malonyl-CoA suivie par la protonation de l'énolate formé, tandis que le domaine  $KS_{\alpha}$  catalyse les réactions de condensation entre un malonate et la chaîne en croissance. Cette dernière est liée au niveau de la  $KS_{\alpha}$ . La sélection des unités d'extension qui sont transférées à l'ACP, se fait par l'action d'une malonyl-CoA : ACP acyltransférase (MCAT). Cependant le gène codant cette dernière enzyme est absent du cluster de gènes codant l'ensemble des PKS de type II. Il a donc été proposé que la MCAT puisse être celle intervenant dans la biosynthèse des acides gras ou encore que les domaines ACP soient capables d'auto-malonylation<sup>16</sup>. Des enzymes post-PKS additionnelles interviennent afin de terminer la formation du polykétide que ce soit des oxygénases, ou des glycosyl- et méthyltransférases.

Figure 8 : Représentation d'une PKS de type II<sup>15</sup>.

### III.2.3 Les PKS de type I

Ce type de PKS est le plus complexe et le plus variable. Il est caractérisé par la présence de plusieurs sites actifs au sein de la même chaîne polypeptidique. Son mode opératoire est de type soit itératif soit modulaire. De même que le type II, il possède une forte homologie en termes de mécanisme et de composition en certains domaines avec les FAS de type I.

Ce type de PKS est composé de modules qui sont eux-mêmes constitués d'un ensemble de domaines minimaux nécessaire à l'incorporation d'un nouveau bloc de construction : AT, ACP et KS. Des domaines optionnels peuvent s'ajouter à cet ensemble de domaines minimaux, pour permettre la modification du bloc de construction : les KR, DH, ER et MT.

#### III.2.3.1 Les PKS de type I itératives

Pour les PKS de type I itératives, l'extension de la chaîne du polykétide se réalise au sein d'un seul module. Une seule multienzyme est donc utilisée à répétition. Au niveau de leurs produits, les PKS itératives permettent la synthèse des acides gras poly-insaturés à longues chaînes (PUFA) retrouvés notamment au niveau des membranes cellulaires de bactéries marines<sup>17</sup>. Les PKS itératives permettent également la synthèse de polykétides de la classe des énediynes chez les bactéries. Ces enzymes présentent une forte homologie de séquence avec celles synthétisant les acides gras poly-insaturés<sup>18</sup>. De plus, parmi le règne des champignons figurent trois catégories de PKS de type I itératives désignées sous le nom de non-réductrices (NR), partiellement réductrices (PR) et hautement réductrices (HR), dépendantes de la composition des multienzymes en domaines de modification<sup>19</sup>.



### III.2.3.2 Les PKS de type I modulaires

Contrairement aux PKS mono-modulaires itératives, les PKS de type I modulaires ajoutent des unités d'extension spécifiques au niveau de différents modules successifs. Ainsi un module, à l'intérieur d'une grande chaîne polypeptidique, ne catalyse la condensation et la modification que d'un acyl-CoA à la chaîne en cours de croissance (Figure 9).

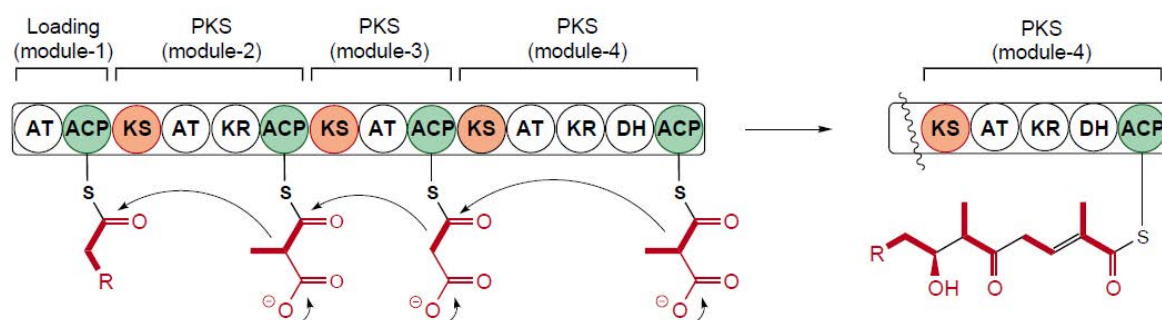


Figure 9 : Représentation d'une PKS de type I modulaire<sup>19</sup>.

De ce fait, cette catégorie de PKS, de par sa composition multimodulaire, est caractérisée par des hauts poids moléculaires pouvant dépasser un million de Dalton. Par exemple, les PKS du système mycolactone MLSA1 et MLSB ont des tailles respectives d'environ 1,8 et 1,4 MDa, et comptent parmi les plus grandes enzymes en biologie<sup>20</sup> (Figure 10).

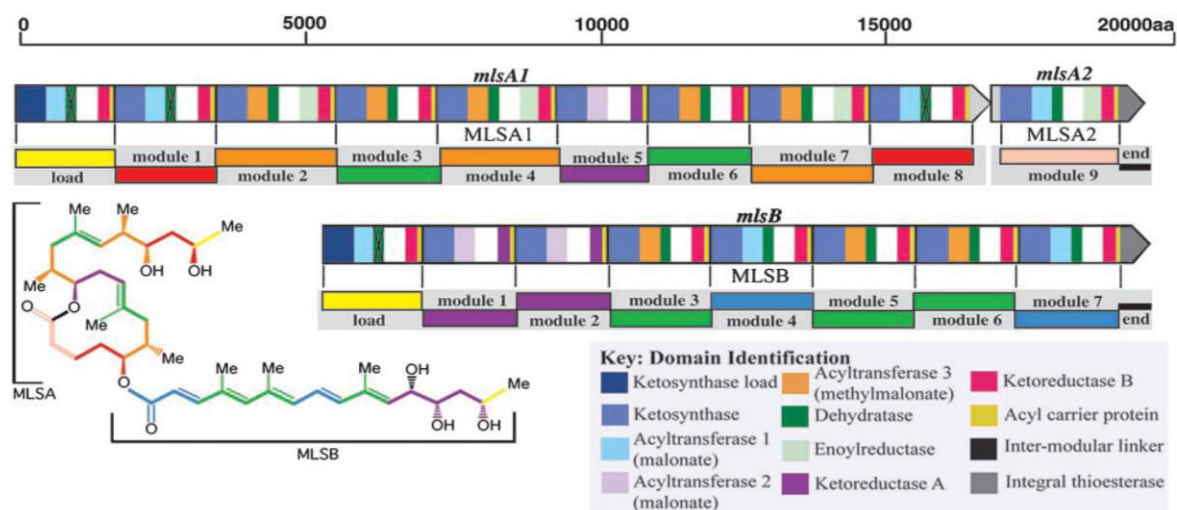


Figure 10 : Représentation des trois polykétide synthases responsables de la formation de la mycolactone<sup>20</sup>. La synthèse de la mycolactone est réalisée par trois PKS codées par trois gènes : *mlsA1*, *mlsA2* et *mlsB*. La première PKS, MLSA1, est composée de neuf modules (un module de départ et huit d'extension). Avec un poids moléculaire de 1,8 MDa, MLSA1 fait partie des plus grandes enzymes rencontrées en biologie.

#### III.2.3.2.1 Le type *cis*-AT

Chez ce type de PKS modulaire, l'ensemble des domaines qui composent un module, que ce soit les domaines minimaux (AT, ACP et KS) ou les domaines

additionnels (KR, DH ou ER) ou encore le domaine de terminaison (TE), est regroupé à l'intérieur de la même chaîne polypeptidique (d'où la désignation *cis*-AT). Commencée à la fin des années 1980, l'étude de la polykétide synthase synthétisant l'érythromycine A (nommée « DEBS ») produite chez *S. erythraea*, a permis de proposer un modèle quant au fonctionnement de ces enzymes de type I modulaires.

Cette PKS s'organise en trois sous-unités : i) DEBS 1, formée du module de chargement et des modules 1 et 2 ; ii) DEBS 2, formée des modules 3 et 4 ; et iii) DEBS 3, formée des modules 5 et 6, ainsi que du domaine thioestérase (TE). A l'intérieur d'un même module, les domaines sont séparés par des séquences de liaison (appelées « linkers »). Ces séquences peuvent être soit des domaines structuraux, soit des séquences de connections flexibles et non-structurées, riches en alanine, proline et résidus chargés, de longueur variable. Elles joueraient un rôle structural important et donc fonctionnel en améliorant les interactions entre les différents domaines et seraient donc essentielles aux mécanismes d'assemblage du polykétide<sup>21</sup>.

La synthèse de l'érythromycine A débute par le recrutement d'un propionate par le module de chargement. La synthèse se poursuit au niveau des six modules suivants qui réalisent l'extension de la chaîne par ajout de groupements propionates dérivés du méthylmalonate. A la fin des six condensations réalisées par les six modules successifs, le domaine TE libère le polykétide non-mature en le cyclisant pour donner le 6-désoxyérythronolide B (6-dEB). Par la suite, le 6-dEB est modifié par 5 enzymes post-PKS (deux hydroxylases, deux glycosyltransférases et une méthyltransférase) pour former l'érythromycine A, la molécule antibiotique<sup>10</sup> (Figure 11).

Que ce soit pour DEBS ou plusieurs PKS de type *cis*-AT, une corrélation est observée entre la composition en domaines actifs des modules de la PKS et la structure du polykétide biosynthétisé. Cette observation est désignée sous le terme de règle de colinéarité. Elle met en avant le caractère modulaire et la logique de biosynthèse de ce type de PKS qui produit le métabolite sous le principe de la chaîne d'assemblage<sup>22</sup>. Cette théorie conduit à la prédiction du produit formé par une PKS grâce à sa composition en modules et elle se révèle être un outil certain pour la synthèse de nouvelles biomolécules *via* la manipulation des gènes codants les PKS. A ce jour, de nouveaux polykétides ont pu être créés par la modification de domaines, de modules et de sous-unités<sup>12</sup>.

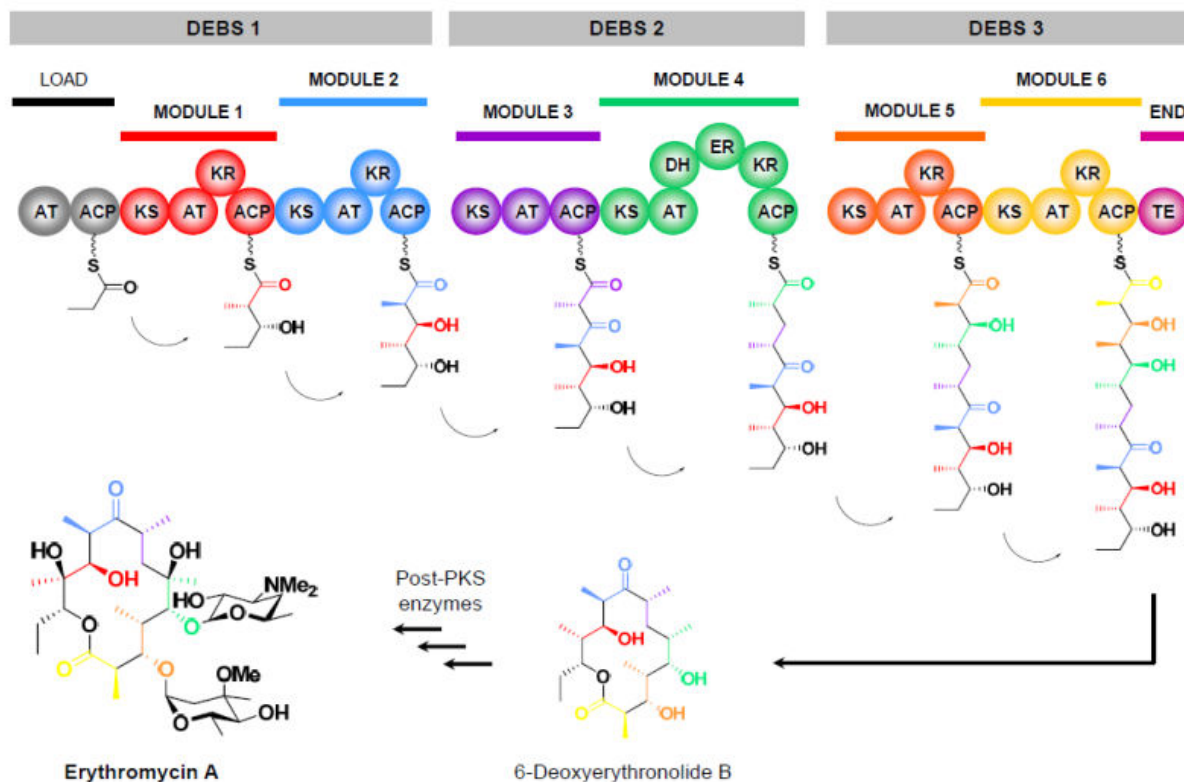


Figure 11 : Représentation de la PKS de l'érythromycine A<sup>10</sup> (adapté par KJ Weissman).

### III.2.3.2.2 Le type *trans*-AT

Pour ce type de PKS modulaire, l'ensemble des domaines d'un même module ne se situe pas exclusivement à l'intérieur d'une même chaîne polypeptidique. Au niveau de ces systèmes, le domaine AT, pourtant décrit comme domaine essentiel, y est absent. Les ACP sont alors acylés par l'intermédiaire de domaines AT discrets, situés en dehors de la PKS. Ce type de PKS a été découvert pour la première fois en 1993 pour la PKS biosynthétisant le bacillaene mais aucune trace de polykétide n'avait été détectée ne pouvant donc pas prouver que la PKS était bien fonctionnelle<sup>23</sup>. C'est seulement en 2002 que la première PKS *trans*-AT, avec un polykétide identifiable, a été caractérisée. Il s'agit de la PKS biosynthétisant la pédérine, une toxine, chez le coléoptère *Paederus*<sup>24</sup>. C'est ainsi que l'hypothèse de l'existence d'un domaine AT discret agissant en *trans* a été émise. Cette hypothèse sera par la suite confirmée par la découverte d'un gène codant un domaine AT à l'intérieur du cluster de gènes responsable de la biosynthèse de la leinamycine. L'étude de cette PKS a notamment aidé à la compréhension du mécanisme des PKS de type *trans*-AT. Le domaine AT discret a été qualifié d'indispensable à la biosynthèse du polykétide et responsable de l'acylation des ACP de chaque module constituant la PKS et ce de manière itérative, *via*

le chargement de malonate. Le fait que le domaine AT agisse de manière itérative induit donc que les différents domaines ACP de la PKS seront acylés par la même unité d'extension. Ce type de PKS se définirait ainsi comme un hybride de PKS de type I et de type II. Afin de diversifier ces unités d'extension, des domaines MT permettent la méthylation du centre  $\alpha$  de l'intermédiaire nouvellement formé par transfert de méthyle de la *S*-adénosylméthionine (SAM)<sup>25</sup>.

Ce type de PKS a également la particularité de présenter une plus grande diversité au niveau de ces modules que ce soit au niveau de l'ordre des domaines, la répétition de domaines, la présence de nouveaux types de domaines, la présence de modules de synthèse de peptides non ribosomiques (NRPS) ou encore de modules découpés sur plusieurs chaînes polypeptidiques<sup>26</sup>.

Un exemple de polykétide synthase de type *trans*-AT est celui de la PKS synthétisant la virginiamycine M chez *S. virginiae*. Ce polykétide est classé parmi les streptogramines de type A. C'est un antibiotique qui inhibe la synthèse peptidique en empêchant la fixation des aminoacyl-ARNt au ribosome en se liant au niveau des sites A et P de sa sous-unité 50S. *S. virginiae* est également responsable de la biosynthèse d'un autre antibiotique, la virginiamycine S. Ce dernier possède un autre mode d'action et fait partie des streptogramines de type B. Il va également se fixer au niveau des sites A et P du ribosome mais en empêchant cette fois-ci la création de la liaison peptidique. L'utilisation synergique de ces deux types d'antibiotique permet d'augmenter l'effet bactériostatique de chaque type de virginiamycine pris séparément<sup>27</sup>. La virginiamycine S est constituée d'un cycle de six acides aminés alors que la virginiamycine M est un macrolide, hybride polykétide/peptide<sup>26</sup>. En effet, la PKS responsable de sa biosynthèse est composée de sept modules PKS et de deux modules NRPS<sup>28</sup>. Les cinq premiers modules sont portés par la chaîne polypeptidique VirA, les modules six et sept par la chaîne polypeptidique VirFG<sup>29</sup> et les modules huit et neuf par la chaîne polypeptidique VirH. Le dixième module qui permet l'ajout de la proline est, quant à lui, déduit puisqu'il n'a pas pu être identifié dans la région génomique séquencée. Une autre particularité de cette PKS réside dans les domaines dupliqués à l'intérieur de plusieurs modules. C'est notamment le cas pour les modules 1 et 5 qui présentent deux domaines ACP et le module 8 qui présente deux domaines *peptidyl carrier protein* (PCP) ainsi que deux domaines hétérocyclases (HC). Ce système comprend également des protéines discrètes. Tout d'abord, le domaine *trans*-AT VirI qui acyle les ACP au niveau des modules PKS, le domaine phosphoantéthéinyl



conduites à partir de 138 modules de PKS provenant de systèmes *trans*-AT. Les différents alignements de séquences ont permis de classer les domaines KS en 16 clades différents. Ainsi, les 138 domaines KS répartis au niveau de ces clades ont été analysés en regard de leur substrat respectif, prédit *via* la structure du polykétide obtenu. Il s'avère que les domaines KS à l'intérieur de chaque clade possèdent la même spécificité de substrat. Par conséquent, l'identité de la structure générée en positions  $\alpha$  et  $\beta$  par le module en amont peut être prédite suivant la catégorisation des domaines KS à l'intérieur des 16 clades. Par exemple, pour l'unité de départ de la synthèse du polykétide, deux types de KS appartenant à deux types de clades différents existent : les KS du clade VI qui ont comme substrat des unités acétyles et les KS du clade XIII qui acceptent des fractions oléfiniques. Ensuite, au niveau des unités d'extension, les domaines KS des clades II, IV et VIII possèdent des substrats hydroxylés en position  $\beta$ , ceux des clades IX, XI et XII des substrats de type oléfinique ou encore ceux du clade VII qui réaliseront la condensation de substrat oléfinique méthylé en position  $\alpha$ . Cette étude, au niveau des domaines KS, permet donc la prédiction de structure du polykétide à partir de la séquence de la PKS.

Au niveau de ces PKS, l'alignement des domaines a également démontré l'existence de clades pour les domaines MT, tandis que pour les domaines ACP, KR et DH, aucun clade n'a pu être identifié. De plus, l'alignement des différents domaines KS à l'intérieur d'un système PKS révèle que l'ensemble de ces domaines ne se retrouve pas au niveau du même clade. Ces observations éclairent l'origine évolutive de ces PKS ; en effet la faible similarité de séquence entre ces domaines KS suggère des phénomènes de transferts horizontaux de gènes entre bactéries<sup>31</sup>. Contrairement aux PKS de type *trans*-AT, celles de type *cis*-AT proviendraient de duplication de gènes et recombinaison entre modules du même système et d'autres PKS au sein de l'hôte<sup>32</sup>. En effet, l'alignement des domaines KS d'une même PKS montre que ceux-ci se situent à l'intérieur d'un même clade. Les deux types de PKS *cis*- et *trans*-AT ont donc évolué de manière convergente en ayant dérivé des FAS (Figure 13).

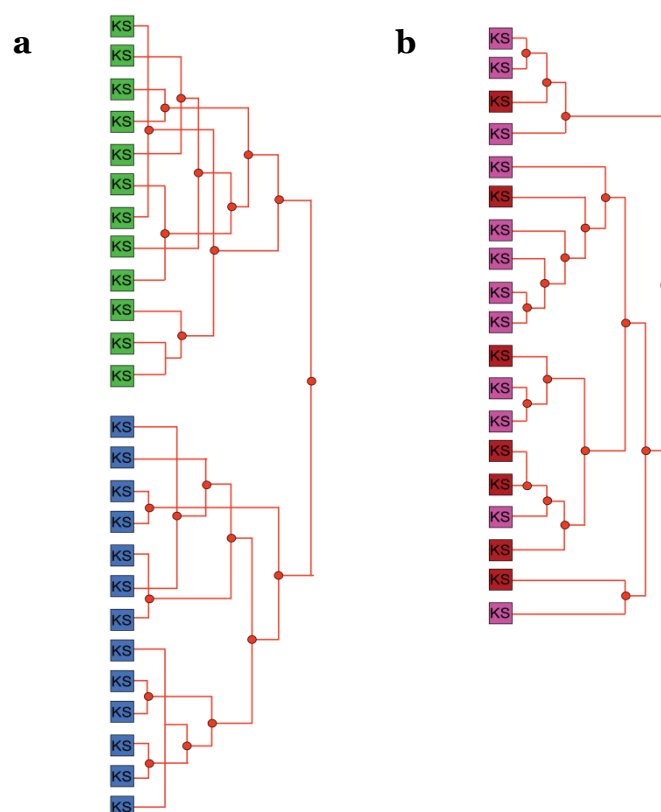


Figure 13 : Représentation des arbres phylogénétiques des domaines KS a) des PKS de type *cis*-AT de l'avermectine (vert) et du macrolide polyène (bleu) et b) des PKS *trans*-AT de myxovirescine (rose) et leinamycine (rouge). Chaque KS de système *cis*-AT est à l'intérieur d'un clade contenant l'ensemble des domaines KS de la PKS, tandis que les KS d'une même PKS de système *trans*-AT sont présentes à l'intérieur de plusieurs clades différents. Ceci met en avant les origines évolutives différentes de ces systèmes<sup>31</sup>.

### III.3 Ingénierie des polykétide synthases

Comme décrit précédemment, le caractère modulaire de certaines PKS, de type I et *cis* ou *trans*-AT, ainsi que la règle de colinarité qui relie intimement la structure du polykétide biosynthétisé avec la composition en domaines actifs des modules de la PKS, font de ces enzymes d'excellentes candidates pour des études d'ingénierie dans le but de produire des analogues de polykétides natifs avec de nouvelles activités.

Plusieurs stratégies ont été mises en place afin de créer une bibliothèque de polykétides non naturels en se basant sur la manipulation des clusters de gènes codant les PKS de type modulaires. Les différentes méthodes employées à ce jour comprennent la mutagenèse, la délétion/inactivation, substitution ou encore l'addition de domaines individuels<sup>33</sup>. La mutagenèse est la méthode la moins drastique qui permet de modifier la sélection des unités de départ et d'extension ainsi que l'état de réduction du polykétide en réalisant respectivement des modifications au niveau des sites actifs des domaines AT et des domaines optionnels<sup>34</sup>. La délétion ainsi que l'addition de domaines permettent de moduler le degré de modification en position  $\alpha$  et  $\beta$  du polykétide et ne concernent qu'un sous-jeu des domaines (KR, DH, ER ou

encore MT pour les PKS de type *trans*-AT). La substitution est appliquée au niveau des domaines AT afin d'altérer le recrutement des unités de départ et d'extension, et donc de modifier le degré de ramification de la chaîne polykétidique, mais également au niveau des domaines KR, DH et ER afin de changer la stéréochimie du produit réduit. L'obtention de produits pour quelques-unes de ces PKS modifiées montre une certaine malléabilité de ces systèmes. Néanmoins, les rendements obtenus *in vivo* sont plus faibles avec les différents mutants que pour la protéine sauvage<sup>5</sup>. Ceci peut être expliqué par plusieurs hypothèses : une modification voire une instabilité structurale engendrée par les changements effectués au niveau de la protéine, une baisse de la réactivité intra-modulaire causée par une diminution de la communication inter-domaines, ou encore une baisse de réactivité des différents domaines vis-à-vis du polykétide modifié en certaines positions contrairement au produit natif.

D'autres stratégies ont également été mises en place pour contrer ces trois problèmes. Par exemple, au lieu d'effectuer des changements au niveau des domaines qui engendrent des problèmes structuraux ou de communication à l'intérieur des modules, l'approche consiste à interchanger/insérer des modules entiers<sup>35,36</sup>. Ce type d'expérience a été réalisé au sein de la PKS DEBS, entre autres. Les PKS recombinantes issues de ces expériences étaient fonctionnelles, et ce notamment grâce à l'optimisation des linkers présents entre chaque module et essentiels à la communication inter-modulaire. Ainsi les modules insérés dans leur contexte non natif présentaient une tolérance envers le substrat en amont qui permet un transfert entre les domaines ACP<sub>n-1</sub> et KS<sub>n</sub>. Cependant cette stratégie présente également ses limites avec de faibles rendements.

Pour pallier ces difficultés, plusieurs points restent à éclaircir sur la biosynthèse des polykétides afin d'avoir recours à une ingénierie efficace. Cinq aspects de la machinerie, que nous étudions au sein de notre laboratoire, doivent être mieux compris : i) comprendre les mécanismes catalytiques de chacun des domaines et résoudre leurs structures seuls ou en complexe ; ii) étudier les interactions inter-domaines essentielles au bon fonctionnement de la catalyse, c'est-à-dire les interactions entre les domaines catalytiques KS, AT, KR, DH, ER et MT et le domaine ACP porteur de la chaîne acyle en cours de synthèse et comprendre comment les domaines agissent ensemble au sein d'un module pour contrôler la stéréochimie du polykétide ; iii) avancer sur la compréhension de la structure modulaire qui révélera l'agencement des domaines entre-eux de manière à appréhender les interactions et



mouvements conformationnels indispensable à la catalyse ; iv) comprendre comment les sous-unités présentes au sein de chaque PKS arrivent à communiquer entre elles de façon efficace et spécifique ; cela implique l'étude des domaines de docking essentiels à ces interactions ; et v) étudier les enzymes post-PKS afin d'établir les relations structure/fonction permettant de mieux employer ces catalyseurs pour diversifier les structures de polykétides ou en tant qu'outils en biologie synthétique. Par la suite, ces cinq points vont être développés.

## IV. Domaines fonctionnels des PKS de type I

Dans cette partie seront présentées les connaissances actuelles sur les structures et mécanismes des domaines retrouvés au sein de ces enzymes.

### IV.1 Les domaines acyltransférase

Les domaines acyltransférase (AT) sont des domaines enzymatiques responsables du recrutement des différents blocs de constructions nécessaires à la formation du polykétide. Ce sont des enzymes monomériques d'environ 300 résidus subdivisées en deux domaines : un domaine catalytique, composé de 240 résidus, qui possède un repliement  $\alpha/\beta$  hydrolase et un domaine *ferredoxin-like* de 60 résidus composé d'une structure  $\beta\alpha\beta\beta\alpha\beta$ <sup>37</sup> (Figure 14).

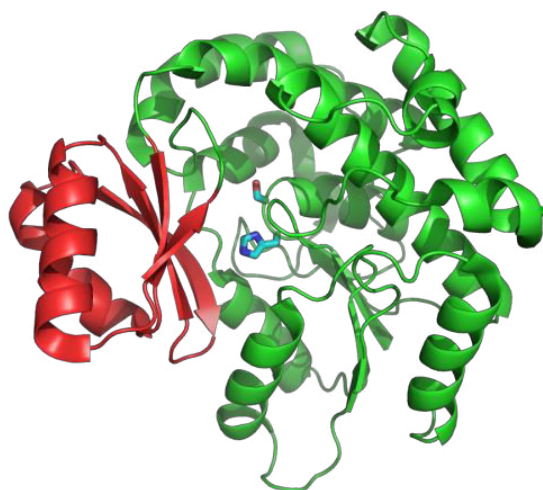


Figure 14 : Structure du domaine AT du module 5 de DEBS<sup>37</sup>. En rouge est représenté le domaine *ferredoxin-like*, en vert le domaine catalytique, et en bleu la sérine catalytique et l'histidine impliquées dans la catalyse.

Au cours de la biosynthèse du polykétide, les domaines AT vont recruter l'unité de départ de biosynthèse du polykétide ou les unités d'extension. Chaque domaine AT

est normalement spécifique d'un seul substrat. Cependant certains domaines AT sont capables de recruter des substrats différents, comme celui du module 12 de la PKS biosynthétisant la stambomycine qui peut charger un 4-méthylhexyl-CoA, un 4-méthylpentyl-CoA, un 5-méthylhexyl-CoA ou un hexylmalonyl-CoA<sup>38</sup>. Ces substrats sont liés à du coenzyme A de manière à les activer afin de faciliter leur recrutement par l'AT en étant un bon groupement partant. Ce recrutement se fait par une réaction d'acylation impliquant un mécanisme catalytique similaire à celui des protéases à sérine. Dans un premier temps, une histidine réalise une catalyse basique générale au niveau de l'hydroxyle d'une sérine afin de l'activer. L'alcoolate ainsi généré fait une attaque nucléophile sur le carbone de la fonction thioester du substrat, conduisant à un état intermédiaire dont la charge négative est stabilisée par un site oxyanion. La charge négative se rabat et provoque le départ du CoA, conduisant à la formation d'un domaine AT lié à son substrat<sup>39</sup> (Figure 15).

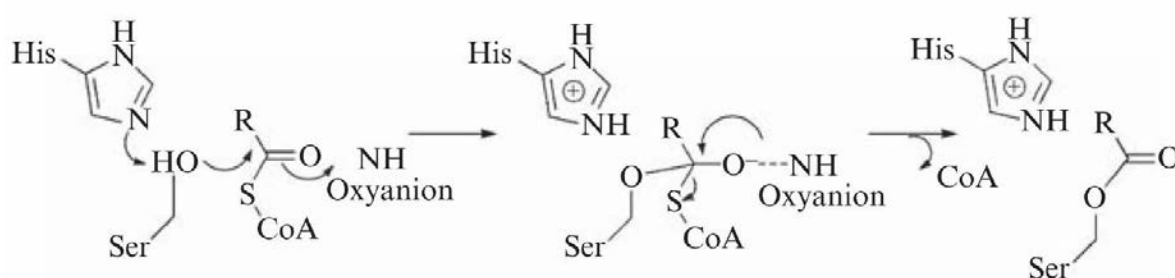


Figure 15 : Représentation de la réaction d'acylation catalysée par le domaine AT<sup>39</sup>.

## IV.2 Les domaines *acyl carrier protein*

Les domaines *acyl carrier protein* (ACP) sont les domaines qui vont servir à porter le substrat au cours des cycles de synthèse sur la PKS ; ils sont comparés au support solide en synthèse chimique. Les ACP sont des petits domaines de 80 résidus dont la structure est composée d'un fagot de quatre hélices  $\alpha$  (Figure 16). La forme active de la protéine est modifiée au niveau d'une sérine spécifique par un bras phosphopantéthéine, conduisant à la forme *holo* de l'ACP. Le substrat sera porté par l'ACP au niveau de ce cofacteur. Ceci amène de la flexibilité et permet ainsi au substrat de pouvoir interagir avec d'autres domaines de la PKS en plus des mouvements de l'ACP à l'intérieur de son module.



Figure 16 : Structure du domaine ACP<sub>5b</sub> du module 5 de la PKS biosynthétisant la virginiamycine M, modifié par la phosphopantéthéine au niveau de la sérine conservée (B Chagot, données non-publiées).

Comme décrit précédemment, au cours de la biosynthèse du polykétide, le domaine AT sélectionne le substrat. Ensuite, il le transfère sur la fonction thiol du bras phosphopantéthéine, en catalysant une attaque du thiolate au niveau du carbone de la fonction ester formée par la sérine catalytique du domaine AT et le substrat. L'oxyanion formé est stabilisé par un site oxyanion. La charge négative se rabat et induit le départ du domaine AT, engendrant la formation d'un ACP acylé<sup>39</sup> (Figure 17).

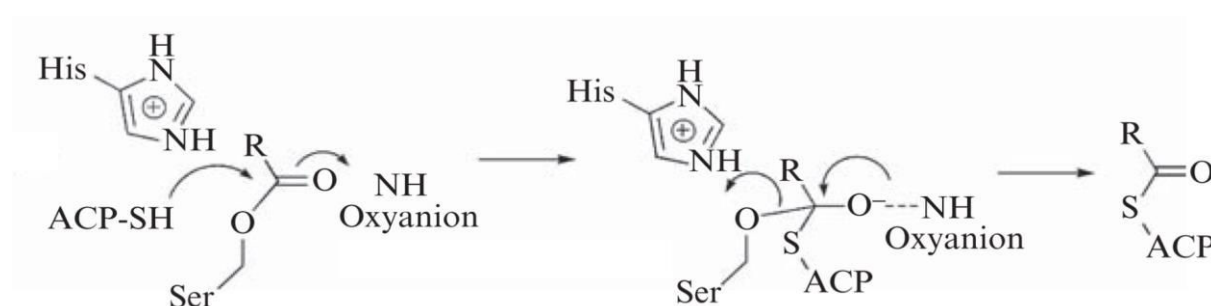


Figure 17 : Représentation de la réaction d'acylation de l'ACP catalysée par le domaine AT<sup>39</sup>.

### IV.3 Les domaines kétosynthase

Les domaines kétosynthase (KS) sont les domaines responsables de la condensation des unités de construction du polykétide entre-elles en catalysant la formation de liaisons carbone-carbone. Ce sont des domaines homodimériques. Un monomère est composé d'environ 430 résidus et possède un repliement de type thiolase composé d'une succession d'hélices  $\alpha$  et de feuillets  $\beta$  ( $\alpha\beta\alpha\beta\alpha$ ) (Figure 18). L'activité enzymatique est assurée par une triade catalytique composée d'une cystéine et de deux histidines<sup>40</sup>.

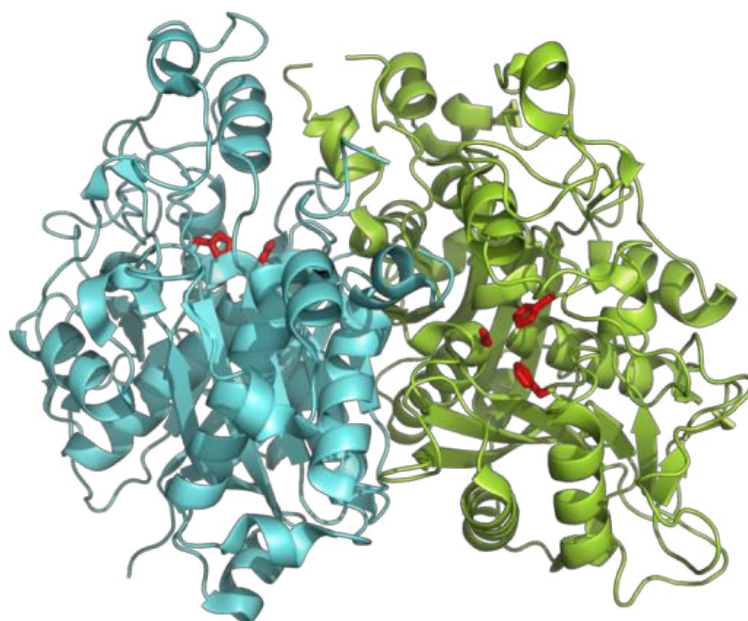


Figure 18 : Structure du dimère de la KS du module 3 de DEBS. En rouge sont indiquées les cystéines catalytiques à l'interface des deux monomères<sup>40</sup>.

Durant la biosynthèse, les domaines KS portent le polykétide en cours de synthèse et réalisent la condensation entre ce dernier et un bloc de construction lié à l'ACP du même module. A ce jour, aucun mécanisme catalytique n'a été étudié sur des domaines KS de PKS. Etant donnée l'homologie entre les PKS et les FAS, le mécanisme proposé est basé sur un domaine KS modèle de FAS<sup>41</sup>. Cette réaction se réalise en deux étapes. En premier lieu, le polykétide lié au domaine ACP du module précédent est transféré au niveau du domaine KS. Cette étape est produite par l'attaque du groupement thiolate de la cystéine catalytique sur le carbone de la fonction thioester formée par le polykétide et la phosphopantéthéine du domaine ACP. La charge négative développée est stabilisée par un site oxyanion, formé par la chaîne polypeptidique. Ensuite, les électrons se rabattent et il y a départ de l'ACP phosphopantéthéinylé qui engendre la formation d'un domaine KS portant le polykétide en croissance. Durant cette première étape, seule la cystéine catalytique intervient ; les deux histidines ne sont pas impliquées (Figure 19). L'attaque nucléophile de la cystéine catalytique est permise par l'abaissement du  $pK_a$  de son groupement thiolate grâce à la proximité du macrodipôle d'une hélice  $\alpha$ .

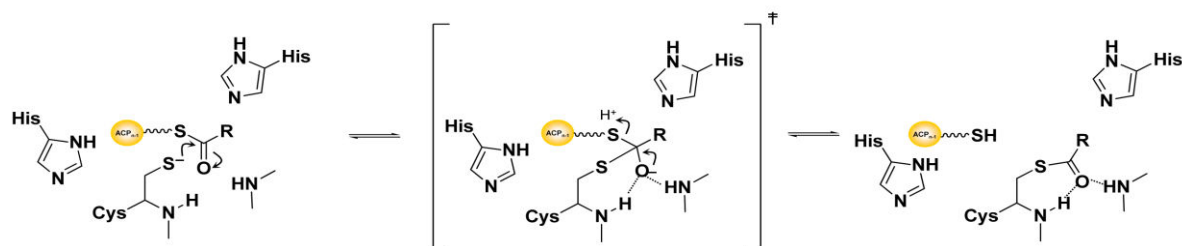


Figure 19 : Première étape du mécanisme de condensation du domaine KS. Le polykétide en cours de croissance est transféré de l'ACP au domaine KS<sup>41</sup> (adapté par KJ Weissman).

Ensuite, le domaine KS réalise la réaction de condensation entre le polykétide et l'unité d'élongation recrutée. Cette réaction se déroule en deux étapes. La première est la décarboxylation de l'unité d'extension. Elle implique une attaque nucléophile par une molécule d'eau activée par catalyse basique générale *via* une histidine de la triade catalytique, générant un intermédiaire tétraédrique qui est stabilisé par la deuxième His. Les électrons se rabattent et ensuite il y a catalyse acide générale de cette même histidine. La deuxième étape est la formation de la liaison carbone-carbone qui débute par la catalyse basique générale d'une histidine pour activer l'énolate. Enfin, les électrons se rabattent, ce qui permet de libérer le produit nouvellement condensé du domaine KS<sup>41</sup> (Figure 20).

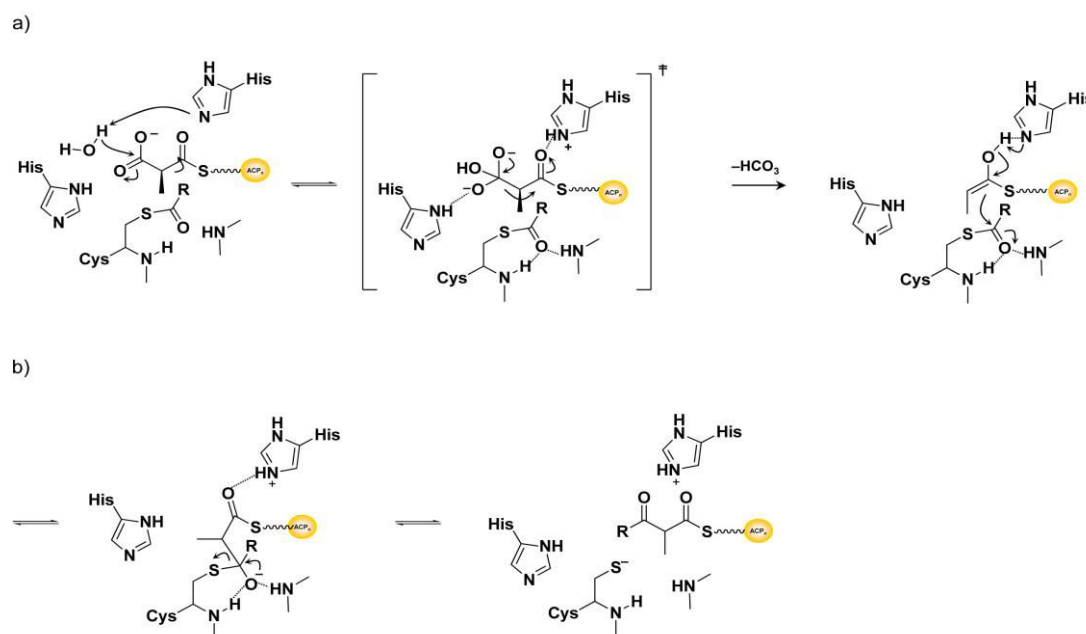


Figure 20 : Réaction de condensation entre le polykétide en cours de synthèse et une unité d'extension. Celle-ci se déroule en deux étapes : a) décarboxylation ; et b) formation de la liaison carbone-carbone<sup>41</sup> (adapté par KJ Weissman).

#### IV.4 Les domaines k tor ductase

Les domaines k tor ductase (KR) sont responsables de la r duction de la fonction c tone du polyk tide en position  $\beta$  en cours de synth se. Ce sont des domaines monom riques compos s d'environ 450 r sidus. Ils se constituent de deux sous-domaines : un domaine catalytique qui comprend 250 r sidus et un domaine structural compos  de 200 r sidus. Ces deux domaines poss dent un repliement de type Rossmann-fold<sup>42</sup> (Figure 21), et donc le domaine global KR pr sente un aspect de pseudo-homodim re.

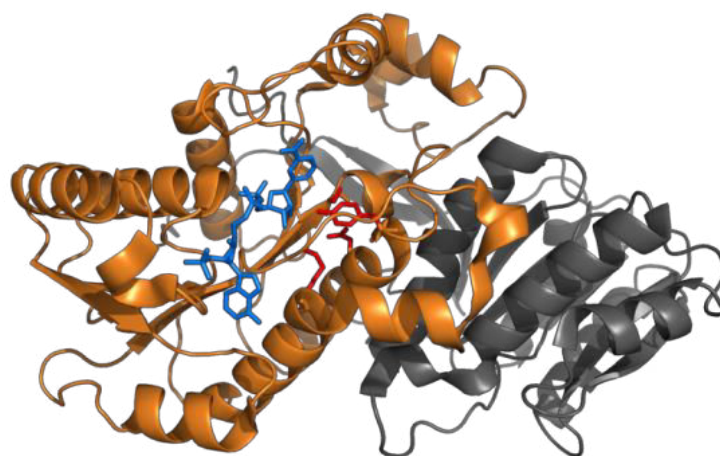


Figure 21 : Structure de la KR du module 1 de DEBS. Elle est compos e de deux domaines : le domaine structural en gris et le domaine catalytique en orange. En rouge sont repr sent s les r sidus impliqu s dans la catalyse et en bleu le NADPH<sup>43</sup>.

Les KR ont pour r le de modifier le polyk tide en cours de synth se au niveau de la position  $\beta$  et de la position  $\alpha$  dans certains cas  galement. Le domaine r duit ainsi la fonction c tone pr sente en position  $\beta$  par l'interm diaire d'un relai proton orchestr  par quatre r sidus catalytiques et le cofacteur r duit NADPH<sup>44</sup>. La r duction se fait par transfert d'hydrure *4-pro-S* du cofacteur sur la fonction c tone suivi par une catalyse acide g n rale de la tyrosine<sup>45</sup>. Ce transfert est facilit  par la s rine et la tyrosine qui augmentent le caract re  lectrophile de la fonction c tone. La tyrosine est stabilis e par la lysine du site actif qui est en interaction avec une mol cule d'eau positionn e par une asparagine du site actif. La catalyse acide g n rale de la tyrosine induit un relai de proton entre ces r sidus et des mol cules d'eau (Figure 22).

Les domaines KR sont  galement responsables de l' pim risation au niveau du carbone  $\alpha$  du polyk tide en cours de synth se<sup>46</sup>. N anmoins, cette caract ristique n'est pas commune   toutes les KR. Cependant   l'heure actuelle, aucun r sidu n'a  t 

identifié comme résidu catalytique conservé. Durant ma thèse, un des objectifs a été d'obtenir la structure d'une KR épimérisante en complexe avec le substrat et le cofacteur afin d'identifier le ou les acides aminés impliqué(s) dans la réaction d'épimérisation.

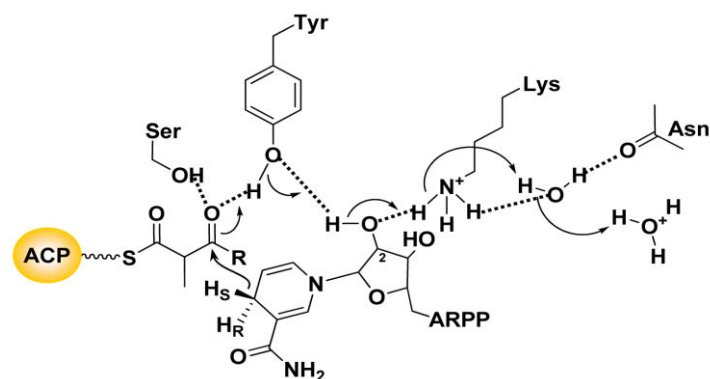


Figure 22 : Réaction de réduction de la fonction cétone du polykétide, lié à l'ACP, par la KR. Elle est réalisée par un transfert de proton faisant intervenir quatre résidus au sein du site actif et le cofacteur NADPH (positionné derrière le substrat)<sup>43-44</sup> (adapté par KJ Weissman).

## IV.5 Les domaines déshydratase

Après la réduction de la fonction cétone en hydroxyle en position  $\beta$  par le domaine KR, le polykétide lié à l'ACP peut subir une réaction de déshydratation si le module comprend un domaine déshydratase (DH). Ceci conduira à la formation d'une double liaison entre les positions  $\alpha$  et  $\beta$  et à la perte de l'hydroxyle en position  $\beta$ . Les domaines DH se composent de 280 résidus et ont un repliement double hot-dog (deux hélices  $\alpha$  entourées chacune de feuillets  $\beta$ ) homodimérique<sup>47</sup> (Figure 23).



Figure 23 : Structure d'un monomère de la DH du module 4 de DEBS composée de deux hélices  $\alpha$  entourées de feuillets  $\beta$  dont le repliement est appelé double hot-dog où l'hélice est la saucisse et les feuillets le pain. En rouge sont indiqués les résidus impliqués dans la catalyse : un aspartate et une histidine<sup>47</sup>.

La réaction de déshydratation est supposée se faire par catalyse basique générale de l'histidine sur un proton en position  $\alpha$ , conduisant à la formation d'une double liaison et au départ de l'hydroxyle en position  $\beta$  réalisé par l'aspartate, grâce à une catalyse acide générale<sup>47</sup> (Figure 24). Cependant, le mécanisme est sujet à débat. En effet, selon une autre hypothèse, l'acide aspartique ne serait pas impliqué dans la catalyse acide<sup>48</sup>.

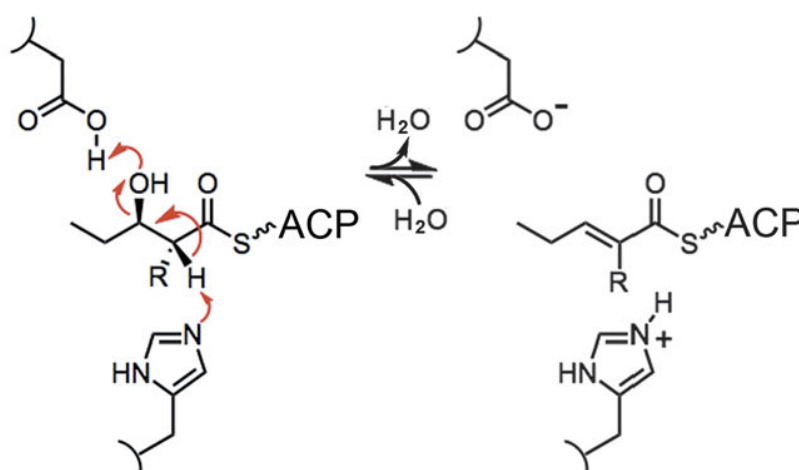


Figure 24 : Proposition du mécanisme de la réaction de déshydratation catalysée par le domaine DH<sup>47</sup>. Cette étape intervient après la modification du polykétide par le domaine KR. L'histidine catalytique arrache un proton en position  $\alpha$ , formant une double liaison et induisant le départ d'une molécule d'eau de la position  $\beta$  catalysé par l'aspartate catalytique.



## IV.6 Les domaines énoyl réductase

Les domaines énoyl réductase (ER) sont les domaines qui vont réduire la double liaison générée par les domaines DH. Cette réaction se fait par l'intermédiaire du cofacteur du domaine ER, le NADPH. Lors de l'acte catalytique, l'ER réalise le transfert d'hydrure 4-*pro-R* du NADPH sur le carbone  $\beta$  du système insaturé, qui entraîne la formation d'un énolate. L'énolate est ensuite protoné de façon stéréospécifique<sup>49</sup> par une catalyse acide générale d'une tyrosine ou encore une lysine activée par un aspartate, suivant le type de domaine ER<sup>42</sup>, conduisant à une double liaison (Figure 25).

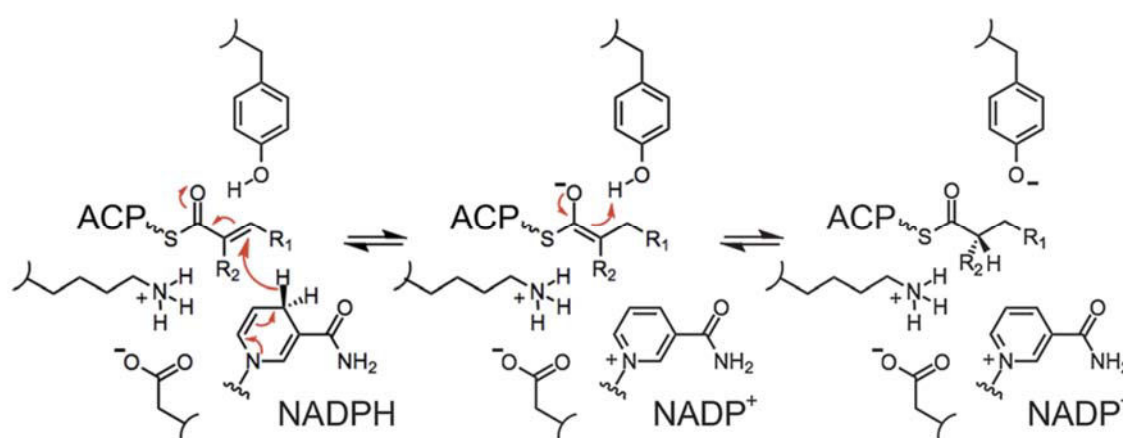


Figure 25 : Réaction de réduction catalysée par le domaine ER. La réduction du système insaturé par transfert d'hydrure du NADPH est suivi par une protonation stéréospécifique qui implique une Tyr (catalyse acide générale) (ou bien une lysine activée par un aspartate selon l'ER)<sup>42</sup>.

Les ER sont des domaines de 310 résidus, dimériques au sein des FAS et monomériques pour les PKS<sup>50</sup>. Ils sont composés de deux sous-domaines : celui de fixation du substrat qui compte 170 résidus et celui faisant 140 résidus qui fixe le nucléotide cofacteur NADPH. Le site actif se situe à l'interface des deux sous-domaines (Figure 26).

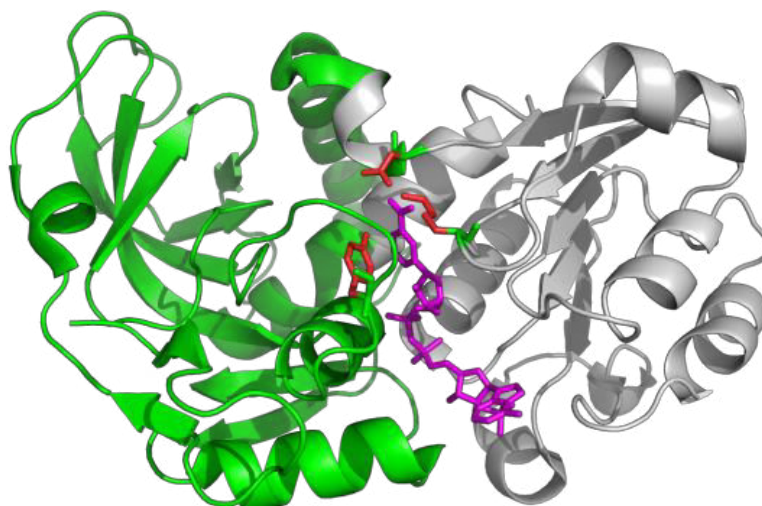


Figure 26 : Représentation de la structure du domaine ER du module 2 de la PKS biosynthétisant la spinosyne chez *S. spinosa*. En vert, le domaine fixant le substrat, et en gris, le domaine fixant le NADPH. En rouge sont représentés les résidus catalytiques : une tyrosine, une lysine et un aspartate<sup>50</sup>.

#### IV.7 Les domaines méthyltransférase

Les domaines méthyltransférase (MT) sont les domaines responsables de la méthylation du carbone en position  $\alpha$  sur un polykétide en cours de croissance et qui possède une cétone en position  $\beta$ . Les domaines MT sont monomériques et comprennent deux sous-domaines : un domaine de 130 résidus fixant le substrat et un autre de 190 résidus fixant le coenzyme SAM. La seule structure de domaine MT résolue est celle de la FAS de mammifères<sup>51</sup> (Figure 27).

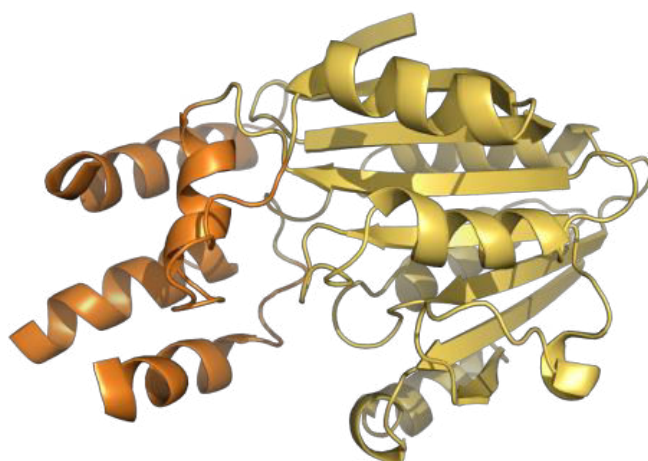


Figure 27 : Représentation du domaine MT de la FAS porcine. En orange est indiqué le domaine fixant le substrat et en jaune le domaine fixant la SAM<sup>51</sup>.

La méthylation en position  $\alpha$  permet d'augmenter la diversité des unités d'extension, notamment au niveau des systèmes de type I itératifs ou dans le cas des

PKS de type *trans*-AT, où l'acyl-CoA sélectionné est toujours le même. Le transfert de méthyle se fait *via* le cofacteur SAM. La réaction commence par une catalyse basique générale au niveau d'un hydrogène lié au carbone  $\alpha$  et la formation d'un énolate. Après attaque nucléophile de l'énolate sur le groupement méthyle de la SAM, le groupement méthyle lié au soufre de la SAM se retrouve en position  $\alpha$ <sup>42</sup> (Figure 28).

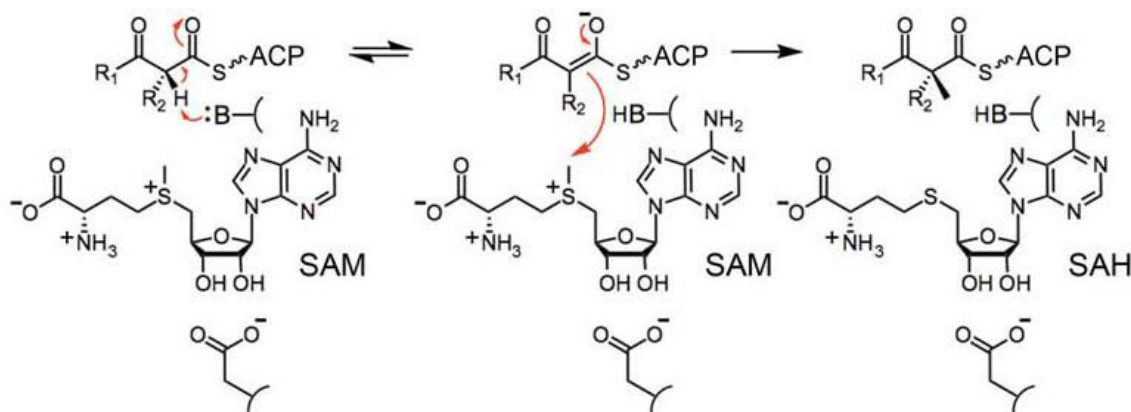


Figure 28 : Réaction de méthylation du polykétide au niveau du carbone  $\alpha$ . La réaction se fait par transfert de méthyle par l'intermédiaire de la SAM. Un catalyseur basique général est nécessaire au déroulement de la réaction dont l'identité est inconnue<sup>42</sup>.

#### IV.8 Les domaines thioestérase

Les domaines thioestérase (TE) sont les domaines qui vont libérer le polykétide de la PKS une fois la synthèse de celui-ci terminée. Ils effectuent cette réaction par hydrolyse ou plus souvent, par cyclisation en employant un nucléophile interne du polykétide. Les domaines TE sont composés d'environ 250 résidus et sont homodimériques. Ils possèdent un cœur  $\alpha/\beta$  hydrolase composé de huit feuillets  $\beta$  et six hélices  $\alpha$  qui est le centre catalytique, et des boucles en périphérie permettant la liaison du substrat<sup>52</sup> (Figure 29).

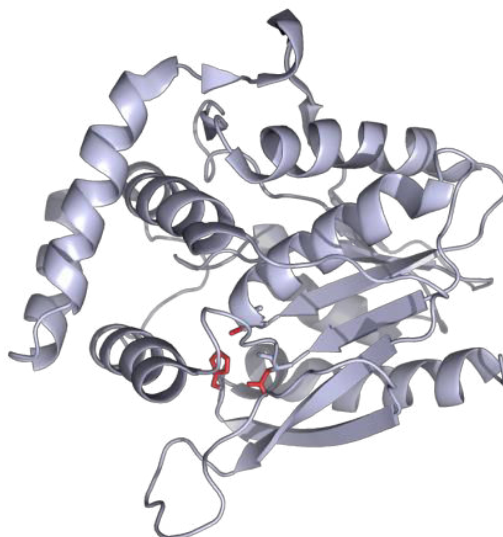


Figure 29 : Représentation du domaine TE de DEBS. La protéine possède un repliement  $\alpha/\beta$  hydrolase. Elle présente une triade catalytique en rouge composée d'un aspartate, d'une histidine et d'une sérine<sup>52</sup>.

Le mécanisme de libération de la chaîne est similaire à celui des protéases à sérine. Le polykétide linéaire, en fin de synthèse, est lié à l'ACP. L'histidine de la triade catalytique active la sérine par catalyse basique générale et celle-ci attaque la fonction thioester. Le polykétide linéaire se retrouve ainsi lié à la sérine catalytique du domaine TE. L'histidine réalise à nouveau une catalyse basique générale au niveau soit de l'hydroxyle en bout de chaîne, soit au niveau d'une molécule d'eau, et les nucléophiles ainsi générés font une attaque sur la fonction ester formée entre le polykétide et la sérine. Ces réactions conduisent soit à une macrolactone, soit à une chaîne linéaire par formation d'un acide carboxylique<sup>42</sup> (Figure 30).

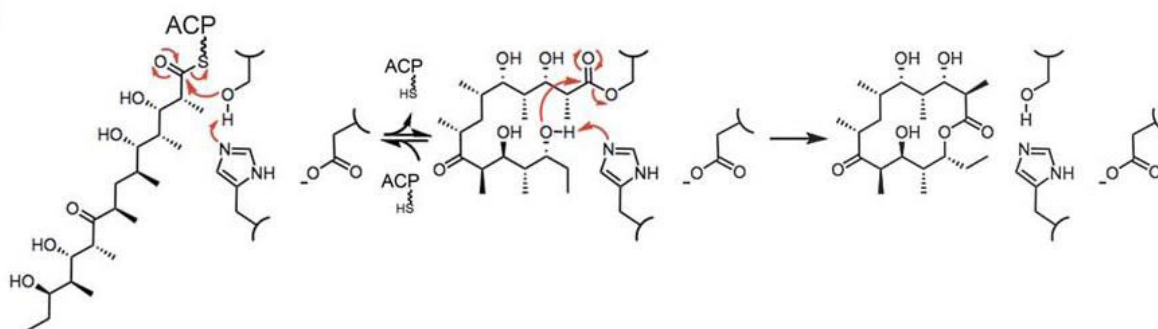


Figure 30 : Représentation du mécanisme de libération du polykétide par cyclisation catalysée par le domaine TE. La réaction se déroule en deux étapes : l'acylation de la sérine avec le domaine TE et la cyclisation *via* l'attaque de la fonction alcool terminale<sup>42</sup>.

## V. Interactions inter-domaines

### V.1 Communications intra-modulaires

Les communications entre les domaines sont indispensables au bon déroulement de la biosynthèse du polykétide. Elles sont essentielles pour le transfert de l'unité de départ ou d'élongation du domaine AT au domaine ACP, mais également pour la condensation du polykétide en cours de fabrication lié au domaine KS avec une nouvelle unité d'élongation attachée à l'ACP. Ces communications entre domaines ACP et KS mais cette fois-ci d'un module  $n$  à  $n+1$  réalisent aussi le transfert du polykétide d'un module à un autre. Les interactions inter-domaines sont importantes aussi pour les processus de modification qui s'opèrent entre le polykétide et les domaines catalytiques MT, KR, DH et ER. Pour les PKS de type *trans*-AT, ces communications s'illustrent aussi par des interactions entre l'ACP et des protéines discrètes intervenant en *trans* dans le processus de  $\beta$ -méthylation, comme pour la PKS de la virginiamycine (Figure 31). Enfin, la libération du polykétide de la PKS *via* le domaine TE nécessite la communication entre le domaine ACP et TE pour le transfert du polykétide linéaire. L'ensemble de ces communications met en avant le rôle central du domaine ACP qui intervient au niveau de chaque étape du processus de fabrication du polykétide (Figure 31).

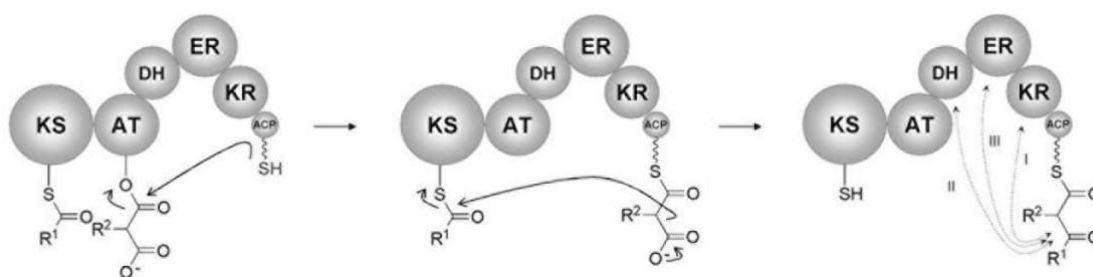


Figure 31 : Représentation des communications inter-domaines au cours d'un cycle d'élongation, mettant en avant le rôle primordial du domaine ACP<sup>53</sup>.

Au vu de l'enchaînement protéique et des contraintes liées à la liaison des différents domaines d'un module, de nombreux changements conformationnels sont essentiels pour que le domaine ACP puisse communiquer avec l'ensemble des autres domaines afin d'assurer la biosynthèse des polykétides<sup>54,55</sup>. Il devient alors primordial de comprendre quels sont les facteurs qui engendrent l'interaction des domaines entre eux et la façon dont les différents domaines du module s'agencent au cours de la

catalyse afin d'optimiser les futures études de biologie synthétique comme par exemple les échanges de domaines.

La première structure, obtenue récemment, d'un module entier a permis de mettre en avant les mouvements des domaines au cours des différentes étapes de la catalyse<sup>54,55</sup>. Ce module est le cinquième de la PKS biosynthétisant la pikromycine et il est composé des domaines KS, AT, KR et ACP<sup>55</sup>. L'acylation du domaine KS par le pentakétide, correspondant au transfert du substrat naturel de l'ACP<sub>4</sub> au KS<sub>5</sub>, s'accompagne de changements conformationnels. Le domaine AT adjacent vient se placer à l'entrée du site actif du domaine KS de manière à éviter une hydrolyse spontanée du substrat. Le mouvement du domaine AT induit un changement de position du domaine ACP. Celui-ci vient alors se plaquer contre le site actif du domaine AT, *via* la formation de ponts salins, afin de faciliter l'acylation de l'ACP par une molécule de méthylmalonate. Néanmoins, l'ensemble de ces mouvements provoqué par l'acylation du domaine KS provoque également la rotation du domaine KR de 180°. Cette rotation est permise par des interactions ioniques entre les domaines AT et KR.

Une fois acylé par l'unité d'extension, le domaine ACP vient se placer contre le site actif du domaine KS pour l'étape de condensation. Lors de la formation de l'hexakétide lié à l'ACP, la forme globale du module est identique à celle du module avec le KS acylé par le pentakétide. Cependant le domaine AT vient complètement bloquer l'entrée du site actif du domaine KS empêchant son acylation par un nouveau pentakétide. L'hexakétide doit ensuite être modifié en position  $\beta$  par le domaine KR. L'ACP se plaque alors contre l'hélice  $\alpha$  et la boucle flanquant le site actif du domaine KR et permet l'entrée du substrat par un côté ou l'autre. Une fois la réduction par le domaine KR, le substrat est entièrement mûré et le domaine ACP se trouve ainsi éjecté de la chambre catalytique afin de transférer l'hexakétide au niveau du domaine KS suivant. Ces études montrent clairement que les interactions de l'ACP avec les différents domaines du module dépendent du substrat acylé sur ce domaine car c'est en principe la fixation du substrat adéquat au sein d'un site actif qui sert à fixer la position de l'ACP au sein du module (Figure 32).

Malgré l'élégance de ce mécanisme, de nombreuses questions restent encore à élucider. Tout d'abord, il faudra vérifier à haute résolution que les différentes positions de l'ACP correspondent à l'entrée du substrat dans les différents sites actifs. Il faudra

confirmer ensuite que ce modèle est identique pour les modules qui comprennent des domaines supplémentaires (MT, DH et ER). Finalement, il faudra observer la manière dont les domaines en *trans* interagissent pour les systèmes *trans*-AT ou encore la manière dont communiquent les domaines ACP et KS localisés sur des modules consécutifs<sup>5</sup>.

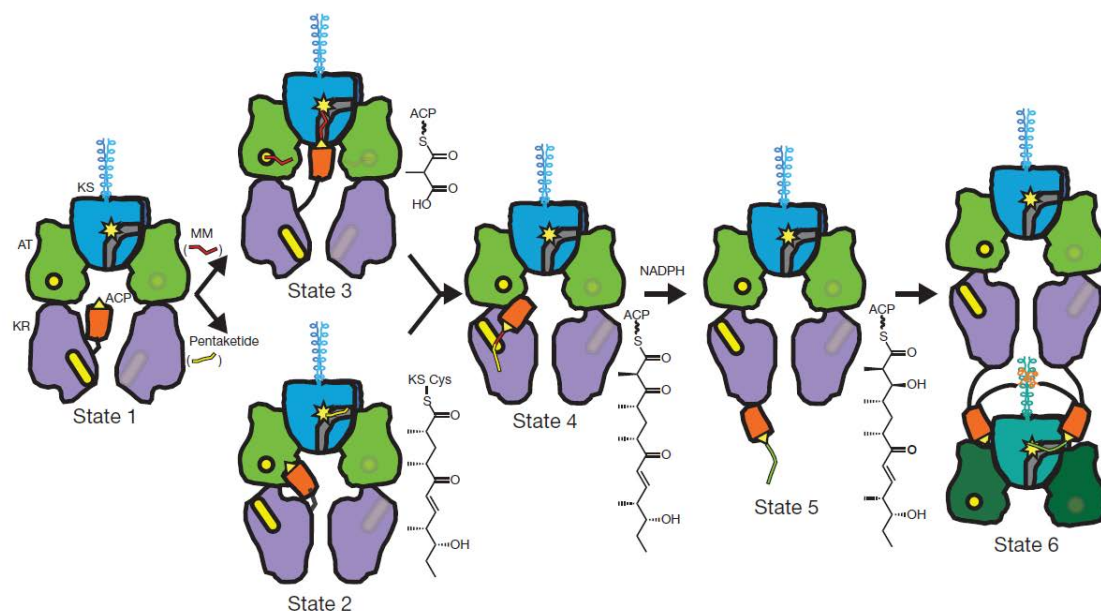


Figure 32 : Représentation schématique des différentes conformations du module 5 de la PKS biosynthétisant la pikromycine.

Elle met en évidence la position de l'ACP aux différentes étapes de la catalyse. Sous forme *holo* non acylé, l'ACP se situe au centre de la chambre catalytique (State 1). Lors de son acylation, celui-ci se plaque contre le domaine AT (State 2) et à l'entrée du site actif de la KS, une fois acylé avec une molécule de méthylmalonate (State 3). Après la formation de l'hexakétide, l'ACP se place à l'entrée du site actif de la KR (State 4). Une fois la réduction en position  $\beta$  effectuée, l'ACP se place hors de la chambre réactionnelle de manière à pouvoir transférer le polykétide au domaine KS du module suivant (States 5 et 6)<sup>55</sup>.

## V.2 Etude du stéréocontrôle

Contrairement à la structure du polykétide, il n'y a pas de corrélation évidente entre la composition en domaines et modules de la PKS et la stéréochimie du polykétide synthétisé, cette dernière étant contrôlée par la PKS. La stéréochimie est une cible attirante de modification car elle est intimement liée à la bioactivité de la molécule. L'importance de la stéréochimie est évidente si on considère, par exemple, la structure du ribosome co-cristallisé en présence de l'érythromycine A<sup>56</sup> (Figure 33). En effet, la configuration des carbones est très importante afin que les groupements hydroxyles ou les oses de l'érythromycine A puissent interagir avec les bases nucléotidiques du ribosome. Ainsi, à chaque étape d'élongation, la stéréochimie est contrôlée que ce soit au niveau de l'hydroxyle du carbone  $\beta$  mais également au niveau du groupement méthyle du carbone  $\alpha$ . Pour preuve, l'érythromycine A possède dix





### V.2.1.2 Les domaines kétosynthase

Les domaines KS quant à eux, sont responsables de la condensation de Claisen avec toujours inversion de configuration aboutissant donc à un centre carbone  $\alpha$  de configuration  $R$ <sup>58</sup> (Figure 34). Cependant, le fait qu'un grand nombre de ces centres méthyles présents dans les polykétides (y compris l'érythromycine A) retrouve une configuration  $S$ , rend compte d'une stéréoinversion après condensation. Ce phénomène résulte d'une activité épimérase. Initialement, il a été suggéré que les domaines KS étaient responsables de cette épimérisation après avoir réalisé l'étape de condensation<sup>59</sup>, mais il a été ensuite démontré clairement que la réaction d'épimérisation est catalysée par les domaines KR<sup>60</sup>.

### V.2.1.3 Les domaines kétoréductase

Les domaines KR sont responsables de la réduction de la fonction cétone en  $\beta$ , aboutissant à des hydroxyles de configuration  $R$  ou  $S$  suivant la face par laquelle la cétone est réduite<sup>45</sup>. De plus, ces domaines agissent comme des épimérase pour modifier la configuration du carbone  $\alpha$ <sup>60</sup> (Figure 34).

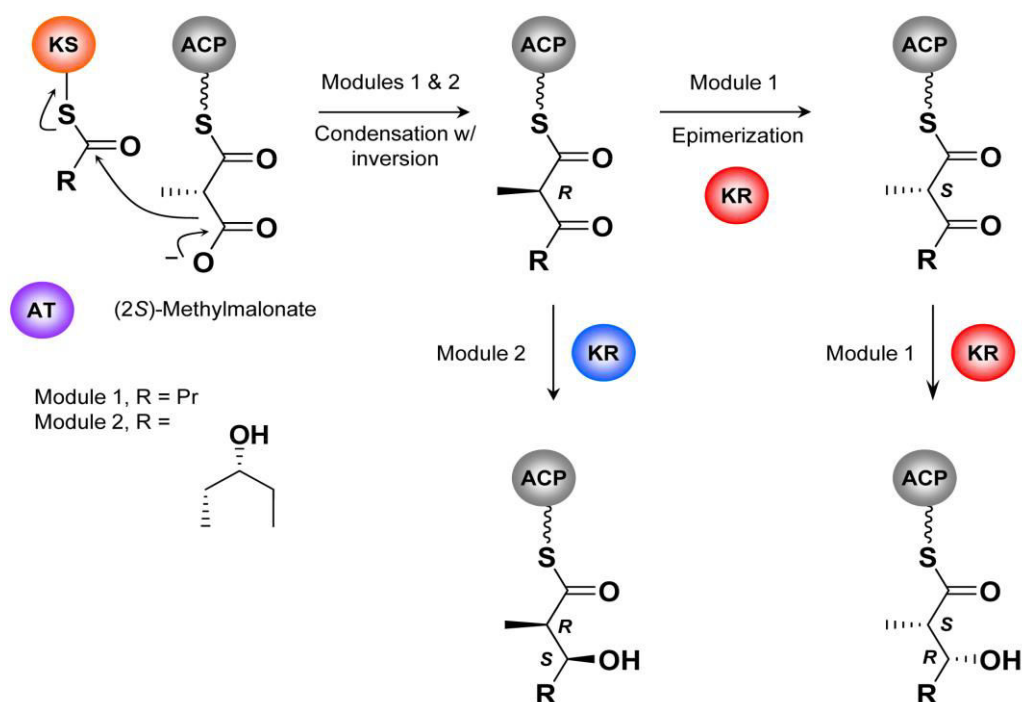


Figure 34 : Domaines candidats au stérocontrôle lors de la biosynthèse de l'érythromycine A (d'après Weissman KJ).

## V.2.2 Les k tor ductases

### V.2.2.1 Activit  r ductase des k tor ductases

Les k tor ductases sont des enzymes appartenant   la famille des *short-chain* d shydrog nase/r ductase (SDR) caract ris es par un repliement de Rossmann (deux unit s  $\beta\alpha\beta$  adjacentes cr ant un domaine  $\beta\alpha\beta\alpha\beta$ ). Ces domaines r alisent la r duction du groupe c to gr ce au transfert de l'hydrure 4'-*pro-S* du NADPH au niveau du carbone  $\beta$  pour donner un hydroxyle<sup>45</sup>. Suivant la direction d'addition de l'hydrure, sur la face *re* ou *si*,   l'int rieur du site actif de la KR, la chiralit  du groupement hydroxyle sera diff rente. La direction de r duction s'av re  tre guid e par l'enzyme de mani re intrins que et donc  tre d pendante du type de KR<sup>61</sup>. De plus, ces enzymes r alisent la production d'un seul diast reoisom re parmi les quatre possibles. Comme la machinerie enzymatique est conserv e entre les KR en termes de r sidus catalytiques et de position du cofacteur, cette st r ospecificit  de r duction d pend de la direction d'entr e du polyk tide dans le site actif, et donc de son agencement par rapport au cofacteur NADPH<sup>60</sup>.

Deux grands types de KR sont alors d crits : le type A, conduisant   la formation d'un hydroxyle de configuration *S*, et le type B amenant   la configuration *R* lorsque les substituants en C<sub>2</sub> sont prioritaires par rapport au C<sub>4</sub> (  noter : la nomenclature de la st r ochimie est invers e lorsque les substituants en C<sub>4</sub> sont prioritaires par rapport au C<sub>2</sub>) (Figure 35). Un troisi me type existe : le type C qui regroupe les KR non actives, incapables de r duire la fonction c tone<sup>62</sup>.

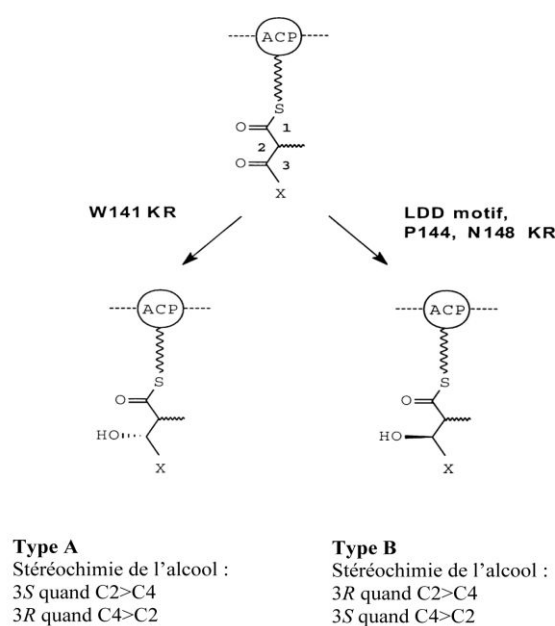


Figure 35 : D termination de la configuration du centre hydroxyle suivant le type de KR<sup>62</sup>.

Les polykétides entrent dans le site actif par le côté gauche pour le type A, alors que cela se fait du côté droit pour le type B. Ces deux types A et B diffèrent notamment au niveau de la séquence en acides aminés de deux motifs : le motif I incluant les résidus 93 à 95 et le motif II incluant les résidus 141 à 148 (numérotés selon un alignement de 68 KR). Ces 2 motifs présentent des résidus conservés, appelés résidus de Caffrey. Le type B présente les résidus conservés Leu<sub>93</sub>, Asp<sub>94</sub>, Asp<sub>95</sub> au niveau du premier motif qui sont absents chez le type A. Le type B possède au niveau du deuxième motif une Pro<sub>144</sub> et une Asn<sub>148</sub> conservées, tandis que le type A ne les possède pas, mais présente un Trp<sub>141</sub> conservé<sup>62</sup> (Figure 36).

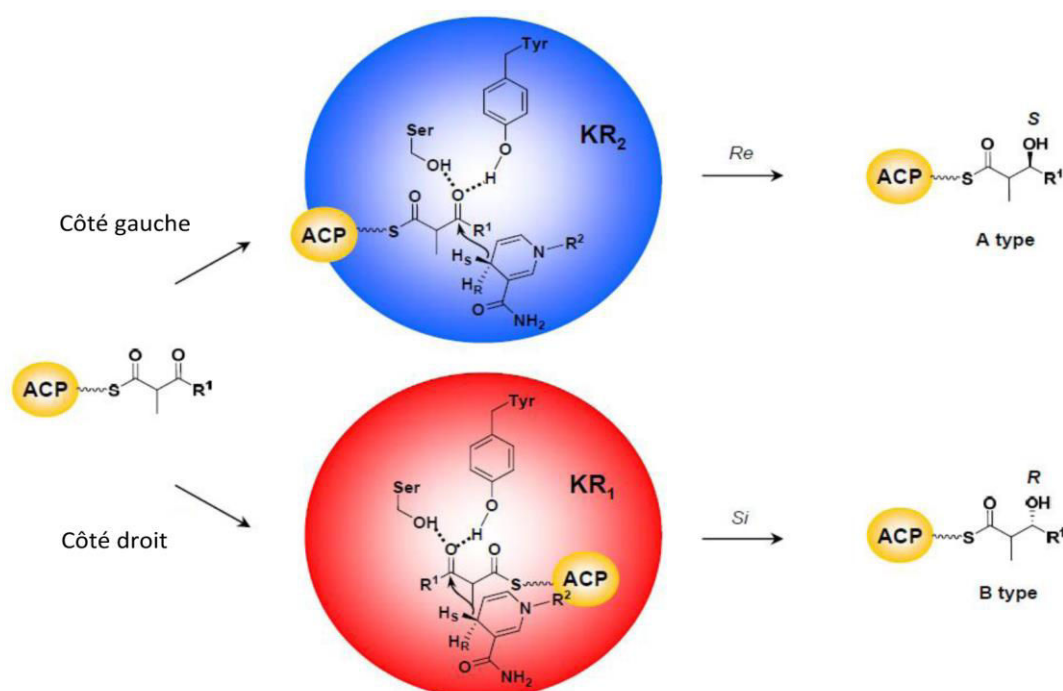


Figure 36 : Stéréospécificité des domaines KR (d'après Weissman KJ). Le cofacteur NADPH est positionné derrière le substrat.

Selon des hypothèses basées sur les structures tridimensionnelles, le Trp<sub>141</sub> guiderait le substrat pour la réduction au sein des KR de type A par contact avec la phosphopantéthéine. Pour le type B, ce serait le motif Leu<sub>93</sub>, Asp<sub>94</sub>, Asp<sub>95</sub> qui serait impliqué, en bloquant l'entrée au site actif par le côté gauche avec l'aide de la boucle flanquante. Ainsi le substrat entrerait par le côté droit grâce à des contacts de la Leu<sub>93</sub> avec la phosphopantéthéine<sup>60</sup> (Figure 37).

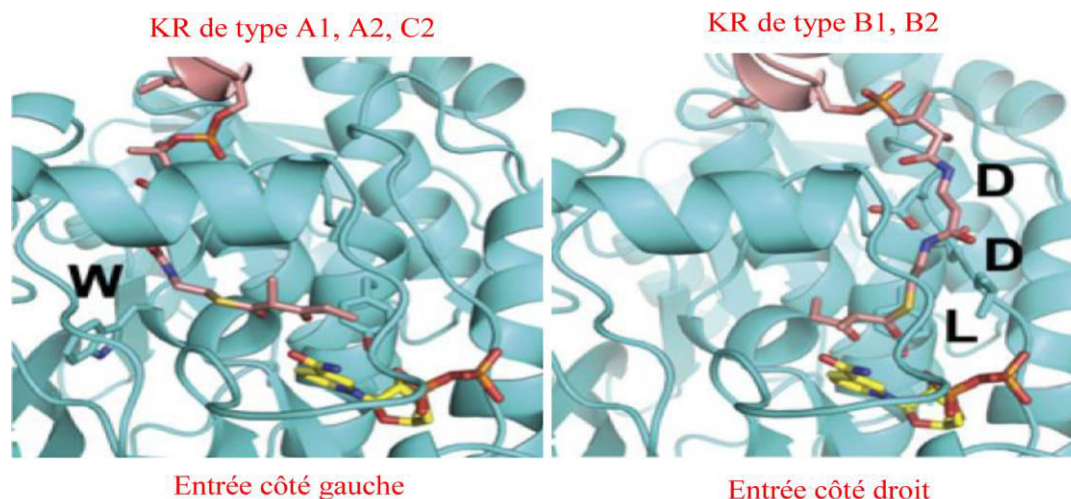


Figure 37 : Modèle du positionnement du groupe prosthétique au niveau du site actif de différents types de KR<sup>60</sup>.

Cependant ces résidus, conservés pour chaque type de KR, ne semblent pas être les seuls responsables de l'entrée du substrat phosphopantéthéinylé dans un sens ou dans l'autre. En effet, des expériences ont été menées sur les domaines KR dans leur contexte natif modulaire ou bien isolés et tendent à confirmer que d'autres entités moléculaires sont impliquées dans la définition du sens de réduction. Par exemple, des études *in vitro*, réalisées avec des domaines purifiés, ont consisté à muter les résidus de Caffrey au niveau de la KR<sub>1</sub> (de type B) et de la KR<sub>2</sub> (de type A), provenant respectivement du premier et du deuxième module de DEBS<sup>63</sup>. Pour eryKR<sub>1</sub>, les mutations simultanées de la Leu<sub>93</sub> et des Asp<sub>94</sub> et Asp<sub>95</sub> au niveau du motif I et de la Phe<sub>141</sub> et de la Pro<sub>144</sub> présents au niveau du motif II ont permis de convertir eryKR<sub>1</sub> en KR de type A, donnant un produit de configuration *S* au lieu de *R*. Pour eryKR<sub>2</sub>, l'introduction des acides aminés Leu<sub>93</sub>, Asp<sub>94</sub> et Asp<sub>95</sub> a provoqué quant à elle l'augmentation du produit naturel de configuration *S* et donc la conservation d'une réduction de type A. Ces résultats ont permis de montrer l'importance de ces résidus conservés et la possibilité d'altérer le stéréocontrôle des KR, mais pas toujours de façon prédite.

Ces mêmes expériences furent reconduites mais avec les domaines KR dans leur contexte naturel, à l'intérieur d'une PKS<sup>64</sup>. Cependant, l'étude de ces multienzymes se révèle être complexe du fait de leur taille. C'est pourquoi l'étude de ces domaines *in vivo* s'est faite au sein d'une petite PKS modèle (appelée DEBS 1-TE) constituée du module de chargement et des deux premiers modules d'extension de la PKS de l'érythromycine A fusionnés avec la TE<sup>10</sup> (Figure 38). Contre toute attente, les cinq mutations effectuées au niveau du site actif d'eryKR<sub>1</sub>, identiques à celles des études *in*

*in vitro*, conduisent à une configuration *R* du groupe hydroxyle, identique à celle du produit sauvage. Pour une raison qui reste à ce jour inconnue, les changements observés dans les études *in vitro*, ne se produisent pas *in vivo*. Dans le cas d'eryKR<sub>2</sub>, l'introduction du motif Leu<sub>93</sub>, Asp<sub>94</sub>, Asp<sub>95</sub> ne change pas, comme dans les études *in vitro*, la réduction native préférentielle des KR de type A. Ces résultats suggèrent au moins la participation de résidus additionnels au sein du site actif dans le contrôle de la stéréochimie.

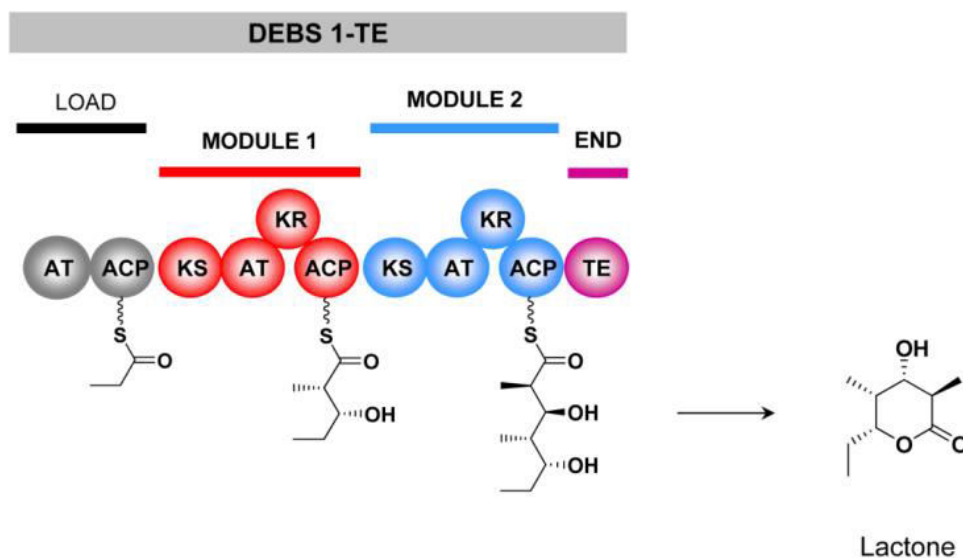


Figure 38 : Représentation de la PKS modèle DEBS 1-TE<sup>10</sup>.

La résolution de différentes structures cristallographiques de KR, a permis d'améliorer la compréhension du mécanisme de réduction. En effet, l'alignement de différentes structures de KR de type A ou B, a permis d'avancer sur la compréhension du sens de réduction et donc d'entrée du substrat à l'intérieur du site actif<sup>65</sup>. Les KR de type A et B présentent une différence d'agencement de leur site actif. Les KR de type A montrent un site actif enclin à la catalyse en présence du cofacteur NADPH contrairement à celles de type B. Ainsi en présence de cofacteur, l'hélice  $\alpha$  et la boucle flanquant l'entrée du site actif des KR et adjacentes au site de liaison du cofacteur, se lient au NADPH par l'intermédiaire d'un réseau de liaisons hydrogène, permettant uniquement l'entrée du substrat par le côté gauche. Pour les KR de type B, et même en présence de NADPH, la boucle est ouverte permettant l'entrée du substrat par le côté gauche ou le côté droit. Néanmoins, l'entrée du substrat entraîne un remaniement du site actif suivi par le réarrangement de l'hélice  $\alpha$  et de la boucle ne permettant l'entrée du substrat que par le côté droit. Les résidus conservés, W et LDD, n'interviennent pas dans le remaniement du site actif des deux KR. En effet, pour les KR de type A, le

tryptophane est trop loin pour participer au positionnement de la boucle contre le cofacteur. Concernant les KR de type B, le motif LDD est assez proche mais son orientation ne permet pas de positionner la boucle correctement. Ceci laisse donc suggérer que d'autres résidus sont impliqués dans le sens de réduction des KR.

L'ensemble de ces observations explique pourquoi le sens de réduction de la fonction cétone est lié aux propriétés intrinsèques de la KR. Ceci est appuyé à nouveau par la structure en microscopie électronique du module 5 de la PKS synthétisant la pikromycine<sup>55</sup>. A partir de la structure à environ 7 Å de résolution, de nouveaux indices ont été obtenus sur le traitement du polykétide à l'intérieur d'un module. Afin de pouvoir modifier le polykétide, le domaine KR se doit apparemment se basculer de 180° pour permettre à l'ACP de se positionner contre l'hélice  $\alpha$  et la boucle flanquant le site actif du domaine KR. Ce basculement est effectué uniquement lorsque l'ACP est sous forme *holo*. A partir de cette position, le polykétide sous forme cétone peut entrer dans le site par un côté ou l'autre. Cependant, une seule configuration est obtenue. Cette observation est donc cohérente avec l'idée que le domaine KR est responsable du sens de réduction (Figure 39).

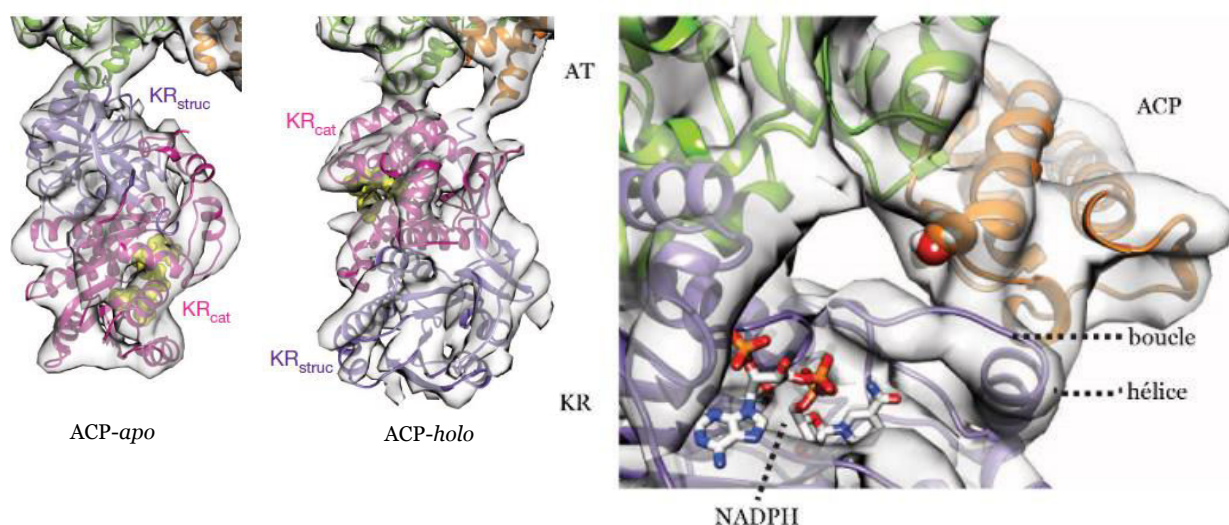


Figure 39 : Représentation de la KR<sub>5</sub> de la PKS pikromycine dans son contexte modulaire. La KR subit un basculement de 180° selon que l'ACP soit sous forme *apo* ou *holo*. L'ACP se positionne alors contre la boucle et l'hélice  $\alpha$  du domaine KR permettant l'entrée du substrat à l'intérieur du site actif de la KR<sup>55</sup>.

### V.2.2.2 Activité épimérase des kétoréductases

Certains domaines kétoréductase sont dotés, en plus de l'activité réductase, d'une activité épimérase<sup>46,60</sup>. Ainsi il faut différencier les KR de type A1 sans activité épimérase et celles de type A2 avec activité épimérase. De la même façon, il faut faire la distinction entre les KR de type B1 et B2. Egalement, le type C se subdivise en deux

sous-types, le type C1 qui ne présente aucune activité enzymatique et le type C2 qui n'est capable que d'épimériser<sup>60</sup>.

Ainsi, des expériences ont été réalisées en échangeant des domaines KR en dehors de leur module d'origine, dans un contexte de modules reconstitués à partir des domaines isolés<sup>66</sup>. De cette façon, l'incubation *in vitro* avec les précurseurs nécessaires pour générer des produits dikétides et trikétides a permis de changer la stéréochimie des centres méthyles et hydroxyles, même de manière incomplète dans certains cas, et de donner un produit de stéréochimie corrélée avec les propriétés intrinsèques du domaine kétoréductase. Le premier module de DEBS, qui est composé d'une KR épimérisante, a été reconstitué et le produit majoritaire obtenu était épimérisé. Le sixième module de DEBS composé d'une KR non-épimérisante a été reconstitué, et le produit majoritaire obtenu était non-épimérisé. Ce même module 6 a été reconstitué ensuite mais en utilisant à la place de la KR de ce module, la KR du premier module. La majorité du produit obtenu était alors épimérisé. De même, le module 1 a été reconstitué mais en utilisant la KR du module 6 à la place de celle du module 1. Le produit majoritaire obtenu était alors non-épimérisé. Ces expériences sont cohérentes avec l'hypothèse selon laquelle la stéréochimie des produits est corrélée avec les propriétés intrinsèques du domaine kétoréductase (Figure 40).

Cependant les bases moléculaires de l'activité épimérase restent encore obscures. Par exemple, le mécanisme de réaction proposé pour chaque type de KR diffère et les résidus impliqués, bien que conservés, ne le sont pas strictement<sup>43,60</sup>. L'espèce proposée comme intervenant dans la catalyse est la tyrosine du site actif ou encore une molécule d'eau correctement positionnée<sup>67</sup>.

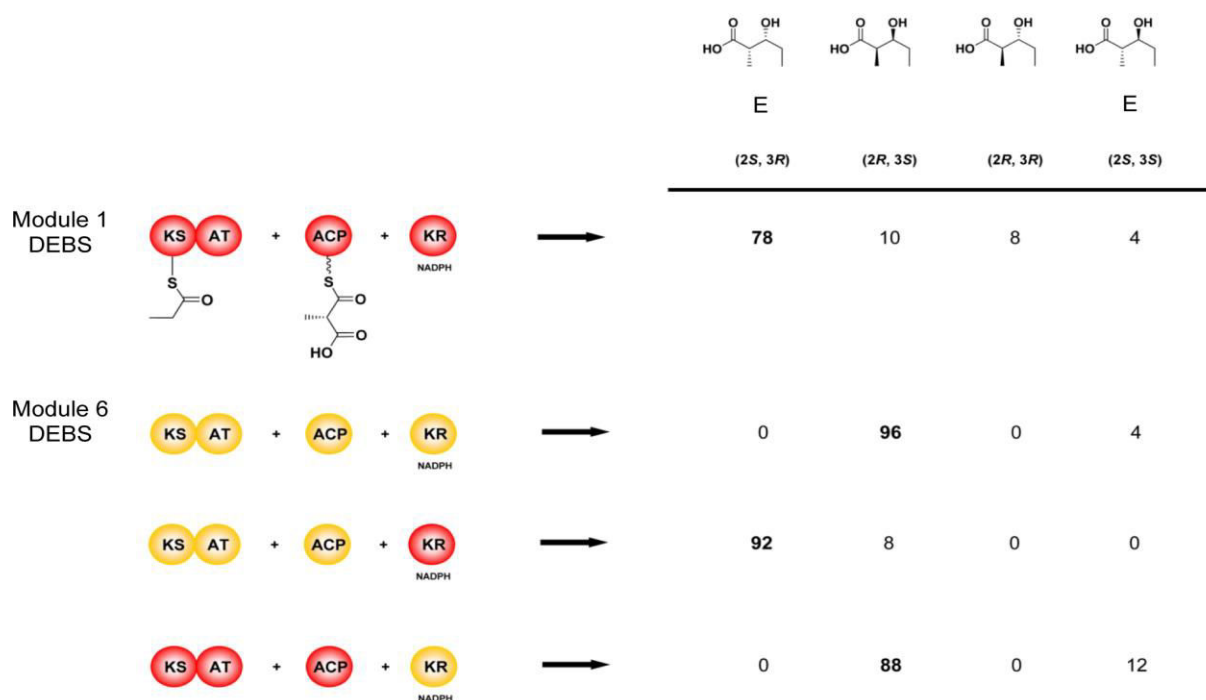


Figure 40 : Reconstitution de modules et analyse de la stéréochimie des dikétides obtenus<sup>66</sup> (adapté par KJ Weissman).

Par ailleurs, des expériences *in vivo* ont consisté à échanger des domaines KR en lieu et place de l'eryKR<sub>2</sub>, une KR non-épimérisante, à l'intérieur de la PKS minimale DEBS 1-TE décrite précédemment, mimant ainsi les expériences fructueuses réalisées *in vitro*<sup>68</sup>. L'insertion des séquences codant les KR s'est faite par l'intermédiaire de plusieurs sites de restriction de manière à moduler leurs extrémités. L'analyse des lactones produites a permis d'appréhender le stéréocontrôle exercé par les différentes KR insérées. Les produits prédits ont été obtenus en insérant des KR sans activité épimérase, de type A1 et B1, comme par exemple la KR du cinquième module de DEBS ou encore la KR du premier module de la PKS de l'avermectine (AVES). Cependant des variations dans les produits obtenus ont été remarquées, suivant les sites de restriction utilisés dans les linkers servant à l'insertion des différents domaines kétoréductase. C'était notamment le cas pour la KR<sub>1</sub> de la PKS de l'avermectine qui délivrait de nouveaux produits minoritaires suivant les sites de restriction utilisés. Ces mêmes expériences ont été réalisées en utilisant des KR de type A2 et B2, respectivement eryKR<sub>1</sub> et la KR du module 7 de la PKS de la rifamycine (RIFS). Toutefois les produits attendus, épimérisés et réduits, n'ont pas été obtenus notamment pour la première KR de DEBS ou alors sous forme de traces comme pour la septième KR de la RIFS PKS (Figure 41).

**Un de mes objectifs de thèse a été d'insérer dans la PKS modèle DEBS 1-TE, l'ensemble des KR épimérisantes (A2 et B2) retrouvées dans des**



**modules ne comprenant qu'un domaine KR en tant que domaine optionnel, de manière à convertir un module non-épimérisant en module épimérisant uniquement par échange de KR.**

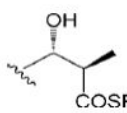
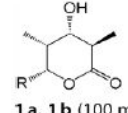
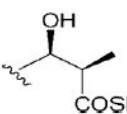
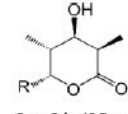
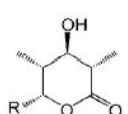
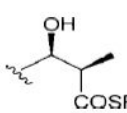
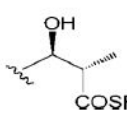
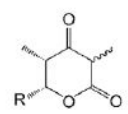
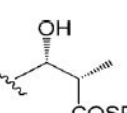
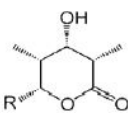
KR insérée (sites)	Type de KR	Fonction normal	Produit majoritaire	Produit minoritaire
DEBS Mod5 ( <i>AvrII-HpaI</i> )	A1		 1a, 1b (100 mg L <sup>-1</sup> )	
AVES Mod1 ( <i>BglII-NheI</i> )	B1		 3a, 3b (25 mg L <sup>-1</sup> )	 4a, 4b (traces)
AVES Mod1 ( <i>PstI-Bsu36I</i> )	B1		3a, 3b (17 mg.L <sup>-1</sup> )	2a, 2b (5,5 mg.L <sup>-1</sup> )
DEBS Mod1 ( <i>AvrII-HpaI</i> )	B2		none	 2a, 2b (traces)
RIFS Mod7 ( <i>SnaBI-NheI</i> )	A2		2a, 2b (27 mg.L <sup>-1</sup> )	 5a, 5b (traces)

Figure 41: Echange de domaines KR à l'intérieur de la PKS modèle DEBS 1-TE et analyse des produits obtenus<sup>68</sup>.

### V.2.3 Stabilisation de la stéréochimie du polykétide

La stéréochimie d'un polykétide en cours de biosynthèse est contrôlée par les domaines KR. Cependant, après la condensation, la fonction  $\beta$ -cétio thioester est très labile et sa configuration restera fixe seulement après la réduction. La stéréochimie doit donc être maintenue puisque le carbone en cette position est sujet à des épimérisations spontanées catalysées par l'eau pouvant altérer la stéréochimie du produit final et donc son activité. Afin d'empêcher ces échanges spontanés avec le milieu, la stéréochimie en C<sub>2</sub> de l'intermédiaire doit être stabilisée avant que la KR n'intervienne. Des études suggèrent que l'ACP est responsable de cette stabilisation<sup>69</sup>. Un didomaine KS-AT acylé par un dikétide a été mis en solution avec un domaine ACP et du méthylmalonyl-CoA. Après condensation et formation d'un ACP acylé par un

trikétide, celui-ci a été isolé et placé en conditions réductrices pour permettre la réduction de la fonction cétone puis en milieu basique pour engendrer l'hydrolyse et la formation d'une trikétide lactone. Les deux produits obtenus ne présentent qu'une seule configuration en C<sub>2</sub>. Un trikétide seul a été réduit et cyclisé mais celui-ci présentait les deux configurations au niveau de ce centre chiral. L'hypothèse de cette étude est donc que le domaine ACP serait bien responsable de la stabilisation du carbone C<sub>2</sub>. Ceci est soutenu par le fait que la constante de vitesse associée à ces phénomènes d'échange est de 0,15 min<sup>-1</sup> pour un polykétide libre dans le solvant et que de 0,01 min<sup>-1</sup> pour un polykétide lié au domaine ACP, en présence du domaine KS<sup>66</sup>. Ces expériences démontrent encore que dans le contexte modulaire, la configuration du carbone est maintenue.

Cependant, l'hypothèse selon laquelle le domaine ACP serait responsable de cette stabilisation a été réfutée par des études ultérieures<sup>70,71,72</sup>. Pour que l'ACP stabilise le carbone en position C<sub>2</sub>, il faut qu'il soit en contact avec ce carbone. Or plusieurs études RMN n'ont montré aucun déplacement de pic entre les spectres [<sup>1</sup>H, <sup>15</sup>N]-HSQC de l'ACP *holo* et l'ACP *holo* acylé avec plusieurs groupements modèles, prouvant l'absence d'interaction entre l'ACP et le carbone C<sub>2</sub>. Ces expériences démontrent qu'il y a donc absence de stabilisation de la configuration en C<sub>2</sub> par le domaine. Il reste donc à déterminer le facteur responsable du maintien de la stéréochimie. **Un de mes objectifs de thèse a donc été de reconstituer le premier module de la PKS de la pikromycine et de confirmer dans un premier temps que le domaine ACP n'est pas responsable de la stabilisation de la position C<sub>2</sub> et ensuite de réaliser des expériences afin d'identifier s'il s'agit du didomaine KS-AT ou du domaine KR.**

## VI. Structure et dynamique des modules

Un des autres points à élucider est la structure des modules et l'agencement des domaines entre eux. Il est donc essentiel de résoudre les structures de différents modules afin de comprendre comment ils se comportent au cours de la catalyse. A ce jour, bien que des structures de domaines isolés aient été résolues aucune structure de module de PKS à haute résolution n'a été publiée. Néanmoins la structure de la FAS de mammifère est disponible. Contrairement aux FAS bactérienne, elle est constituée d'une unique chaîne polypeptidique et non pas de domaines isolés et semble donc être

un bon modèle pour comparer sa structure avec celles des modules de PKS<sup>5173</sup> (Figure 42).



Figure 42 : Représentation de la similarité de composition en domaines entre la FAS de mammifère et les modules de PKS (ici le module 4 de DEBS). Ici la constitution en domaines est identique, sauf un domaine méthyltransférase ( $\psi$ ME) inactif supplémentaire et le domaine TE qui libère l'acide gras fabriqué et présent en Cter du dernier module des PKS<sup>51</sup>.

La FAS et le module 4 de DEBS sont composés des mêmes domaines à l'exception du domaine  $\psi$ ME qui est un domaine méthyltransférase inactif et le domaine TE. Au niveau de la FAS, l'ensemble de ces domaines est séparé par des linkers non structurés à l'exception du linker entre le domaine KS et MAT qui est structuré et constitué de deux hélices  $\alpha$  et trois brins  $\beta$ . La structure de l'enzyme a été obtenue par cristallographie aux rayons X à une résolution de 4,5 Å<sup>73</sup>, puis à une résolution de 3,2 Å. La protéine est homodimérique. Elle est agencée en deux chambres réactionnelles. Chaque chambre est constituée de deux zones : celle responsable de la condensation, comportant les domaines KS et MAT, et l'autre responsable de la modification, qui comporte les domaines  $\psi$ ME, KR, DH et ER. A l'interface des deux parties se trouve un linker central situé entre le domaine MAT et le domaine DH. Ces deux linkers, appartenant à chacun des monomères, se croisent donnant à la structure globale une forme de X. La dimérisation est assurée par l'intermédiaire des domaines KS, DH et ER. A noter que pour les domaines ACP et TE, leurs structures n'ont pas pu être résolues probablement à cause de leur flexibilité. Le fait que le domaine ACP soit flexible lui permet d'aller interagir avec l'ensemble des domaines constituant la FAS. Cette flexibilité provient des linkers situés de part et d'autre du domaine ACP, d'une longueur de 40 Å en Nter et de 80 Å en Cter. Les domaines ACP sont monomériques et sont modélisés comme étant présents dans le centre réactionnel créé par les deux zones de condensation et modification, permettant ainsi aux domaines ACP d'interagir avec l'ensemble des domaines catalytiques (Figure 43).

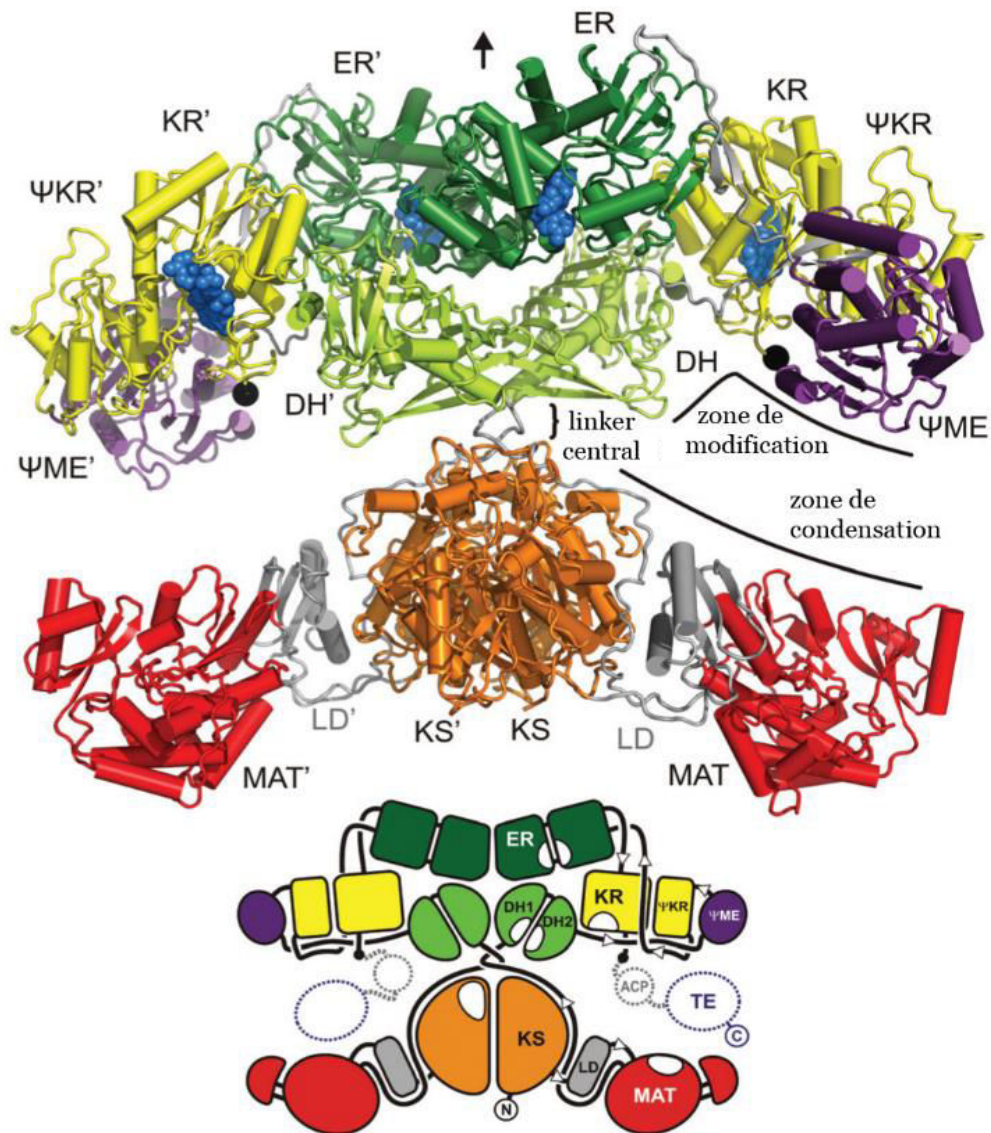


Figure 43 : Représentation de la structure de la FAS de mammifère. L'axe de dimérisation est montré par l'intermédiaire de la flèche. Sur le schéma sont indiquées les positions présumées des domaines ACP et TE<sup>51</sup>.

Cette structure modèle a été utilisée afin de comprendre le fonctionnement des PKS modulaires. Cependant il existe des différences entre ces dernières et les FAS. En effet, les PKS modulaires ne fonctionnent pas de manière itérative à l'image des FAS. De ce fait, la communication inter-modulaire, c'est-à-dire entre l'ACP d'un module  $n$  et le KS du module  $n+1$ , n'existe pas au niveau de ces systèmes. De plus, les PKS possèdent une composition en domaines qui varie suivant les modules. Malgré ces différences, les domaines des FAS possèdent le même repliement que ceux des PKS, qu'il s'agisse des domaines AT<sup>37</sup>, KS<sup>40</sup>, ACP<sup>74</sup>, KR<sup>43</sup> ou encore des domaines DH<sup>47</sup>. Les FAS qui se comporteraient comme des modules de PKS isolés, sont alors utilisées pour construire des modèles de structure de PKS. Cependant quelques différences sont à

noter, décrites précédemment, comme le fait que le domaine TE s'homodimérise au sein des PKS tandis que le domaine ER ne s'homodimérise pas<sup>50</sup>.

Ainsi un modèle de PKS a été créé à partir de l'ensemble des observations faites sur les FAS et les domaines de PKS isolés. Les deux synthases possèdent des structures différentes du fait que le domaine ER soit monomérique chez les PKS et se retrouve ainsi placé à l'extérieur des deux chambres catalytiques. Ceci a pour effet de rendre le domaine ACP plus accessible aux différents domaines en prenant la place du dimère de domaines ER<sup>50</sup>. Cependant, les linkers de part et d'autre de l'ACP suggèrent des changements structuraux significatifs, car leur taille est insuffisante pour qu'il puisse interagir avec le domaine KS et les domaines de modifications de son module et surtout le domaine KS du module suivant (Figure 44).

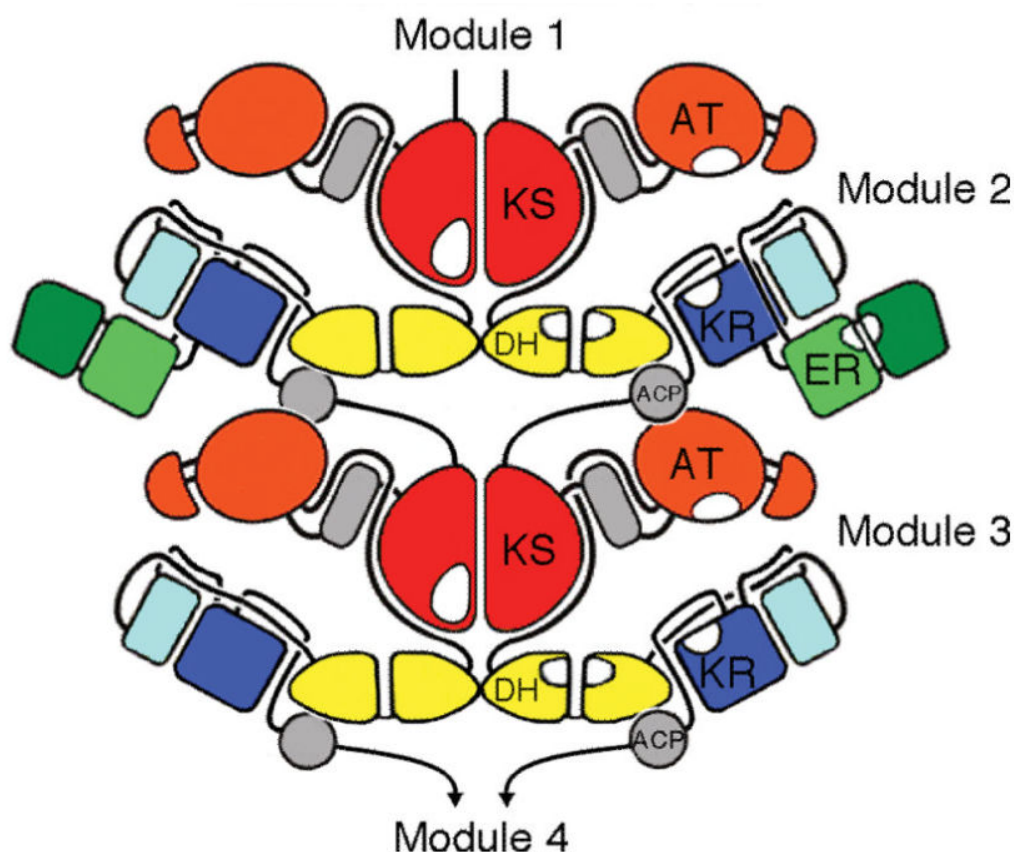


Figure 44 : Représentation d'un modèle de deux modules successifs de PKS basé sur la structure de la FAS de mammifère et des domaines isolés de PKS<sup>50</sup>.

L'obtention de la structure du module 5 de la PKS biosynthétisant la pikromycine par cryo-microscopie électronique a contredit le modèle de la PKS<sup>54</sup>. Le module est un homodimère symétrique avec une conformation en arche où la KS homodimérique se situe au niveau du dôme, les domaines AT sur les côtés et les

domaines KR à la base. Le domaine ACP sous forme *holo* se place, quant à lui, près du domaine AT ou KR (Figure 45). Le linker entre les domaines KS et AT est structuré, servant d'interface entre les deux domaines.

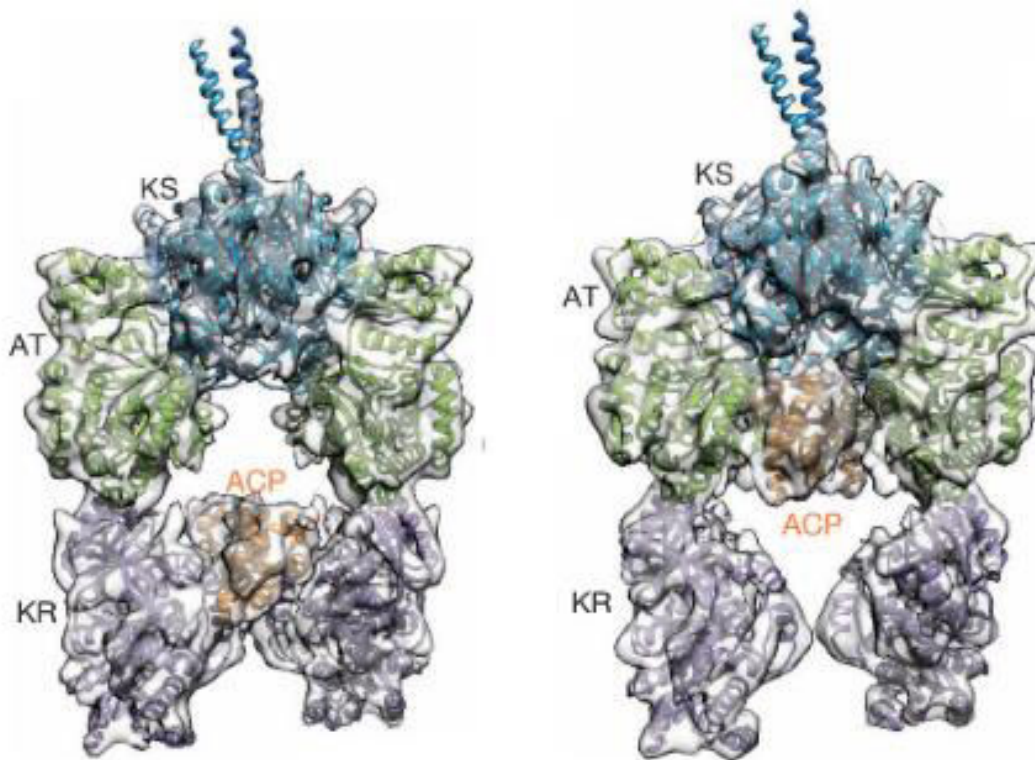


Figure 45 : Structure par cryo-microscopie électronique du module 5 de la PKS biosynthétisant la pikromycine en forme *holo*. Pour le conformère de gauche, l'ACP se situe à proximité de la KR, alors que pour le conformère de droite, l'ACP se situe à proximité du domaine AT<sup>54</sup>.

La forme globale de ce module ne donne ainsi qu'une chambre réactionnelle. Cela signifie que si deux chaînes sont synthétisées simultanément, elles vont croître dans une seule chambre réactionnelle contrairement aux FAS qui sont composées de deux chambres réactionnelles, une par chaîne polypeptidique. L'ensemble des sites actifs des domaines du module ne sont accessibles qu'à l'ACP du module, ceci empêche des interactions aspécifiques avec les ACP d'autres modules, maintenant ainsi l'intégrité du produit final. L'interaction de l'ACP avec l'ensemble des domaines de son module est rendue possible grâce au linker entre le domaine ACP et KR, d'une taille de 43 acides aminés.

Ceci permet entre autre à l'ACP d'interagir avec le domaine KS de son module mais également avec le domaine KS du module suivant. La différence se fait au niveau du site d'interaction. Pour l'ACP du module 5, l'entrée du substrat au niveau du site actif de la KS<sub>5</sub> se situe à l'intérieur de la chambre réactionnelle, tandis que pour l'ACP

du module 4, l'entrée au niveau du site actif de la KS<sub>5</sub> pour réaliser la réaction de transfert du polykétide se fait en dehors de la chambre réactionnelle, au sommet de la structure en arche. Ainsi, il y a une démarcation claire entre les deux ACP en termes de leurs relations à la même KS.

La structure à basse résolution (env. 20 Å) d'un *apo* module de type *trans*-AT a également été obtenu par notre équipe<sup>29</sup>. Le module modèle (module 5 de la PKS virginiamycine) est composé d'un domaine KS ainsi que de deux domaines ACP à la suite. La forme globale de la structure obtenue par SAXS est celle d'une arche de la même manière que les PKS de type *cis*. Le module est homodimérique avec le domaine KS<sub>5</sub> sous forme d'homodimère au sommet de l'arche, les premiers ACP (ACP<sub>5a</sub>) de chaque côté et les deuxièmes ACP (ACP<sub>5b</sub>) à la base. La structure suggère que l'ACP<sub>5a</sub> pourrait servir à la réaction de condensation avec le domaine KS<sub>5</sub> alors que l'ACP<sub>5b</sub> servirait pour la modification de la chaîne par la cassette de  $\beta$ -méthylation du polykétide nouvellement condensé. Le linker situé entre le domaine KS<sub>5</sub> et l'ACP<sub>5a</sub> est structuré ce qui pourrait permettre au domaine AT agissant en *trans* de se « docker » contre le domaine KS<sub>5</sub> de façon similaire à ce qui existe pour les systèmes *cis*. Cependant avec une longueur de 159 résidus, ce linker est compact et maintient l'ACP<sub>5a</sub> proche du domaine KS alors que le linker situé entre les deux ACP est court mais apporte une grande flexibilité au niveau de l'ACP<sub>5b</sub> lui permettant d'interagir avec les domaines de modification (Figure 46). Plusieurs hypothèses restent à vérifier expérimentalement comme par exemple le fait que les deux ACP jouent bien des rôles différents.

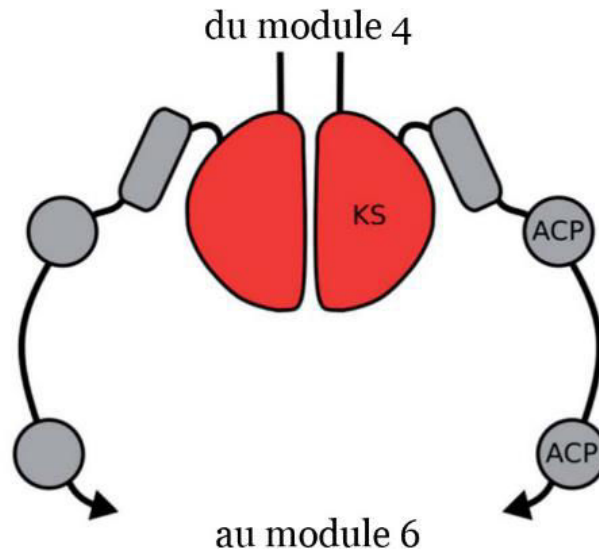


Figure 46 : Modèle du module 5 de la PKS biosynthétisant la virginiamycine, basé sur des données SAXS<sup>29</sup>.

## VII. Les domaines de docking

Une des propriétés communes de quasiment tous les systèmes PKS est que les modules sont distribués entre plusieurs sous-unités. Ainsi, pour que la PKS soit fonctionnelle et conduise à un polykétide de structure précise, il est indispensable que les différentes sous-unités interagissent entre elles et ce de manière spécifique. Pour ce faire, des domaines d'interaction ou de 'docking' (DD) permettent la liaison des différentes chaînes polypeptidiques de la PKS et donc la formation d'une chaîne d'assemblage complète. Ils interviennent dans la reconnaissance des sous-unités de manière spécifique, différenciant alors les sous-unités entre elles en évitant des interactions aspécifiques entre des sous-unités d'une même PKS qui ne se suivent pas ou bien entre deux sous-unités de PKS différentes. Cette reconnaissance engendre donc le transfert spécifique du polykétide en cours de synthèse d'une chaîne polypeptidique à une autre à l'intérieur d'une PKS et permet la synthèse du produit attendu. Un autre type d'interaction existe, ne mettant pas en jeu uniquement les DD. Ainsi, elle est médiée par les DD et également par le domaine ACP, qui entre en interaction avec le domaine KS de la chaîne polypeptidique suivante<sup>75</sup>.

Ces DD sont de petits domaines présents aux extrémités Nter et Cter de l'ensemble des chaînes polypeptidiques des PKS de type *cis*-AT, excepté en Nter de la première chaîne et en Cter de la dernière. Quant aux PKS *trans*-AT, ces domaines ne sont présents qu'à certaines jonctions (KJ Weissman, données non-publiées). Il



devient alors essentiel de caractériser ces domaines de docking, notamment d'un point de vue structural, afin de comprendre la manière dont ils interagissent entre eux. De cette façon, la biosynthèse de nouveaux polykétides pourra être envisagée en modifiant l'ordre des sous-unités des PKS.

Des études antérieures à celles des résolutions de structures des différentes classes de DD existantes ont consisté à construire des PKS hybrides à partir des PKS biosynthétisant l'érythromycine A, la pikromycine et l'oléandomycine<sup>76</sup>. Ces trois polykétides ont des structures proches et sont biosynthétisés par des PKS qui partagent une forte homologie de séquence puisqu'ayant évoluées d'un ancêtre commun. Ainsi, les troisième et quatrième sous-unités de la PKS de la pikromycine ont été remplacées par la troisième sous-unité de la PKS de l'érythromycine A ou de l'oléandomycine. Les PKS résultantes de ces combinaisons étaient fonctionnelles, et présentaient des rendements en polykétide élevé. Ce résultat démontre la capacité intrinsèque des DD à communiquer, car ils n'ont pas été modifiés. L'avantage de l'échange simple de chaîne polypeptidique est de conserver intacte la structuration de la chaîne polypeptidique en intervenant sur les sous-unités, contrairement aux autres méthodes d'ingénierie qui visent à échanger des domaines ou des modules. Néanmoins, l'idée serait de pouvoir combiner plusieurs sous-unités provenant de PKS avec des compositions en domaines et modules bien différentes. Ainsi, de nouvelles études ont consisté à intervertir des DD<sup>77</sup>. L'échange d'un couple de DD par un autre peut conduire à un transfert de polykétide aussi efficace que pour la PKS native comme par exemple, lors de l'échange des DD présents à l'interface des première et deuxième sous-unités de DEBS par ceux à l'interface des deuxième et troisième. Cet échange a conduit à la formation d'un polykétide avec un rendement identique à celui de la protéine sauvage. Par ailleurs, il a été démontré que l'utilisation d'un couple de DD qui n'interagit pas interdit un transfert de polykétide d'une sous-unité à une autre ou bien la PKS résultante montre une activité peu efficace. Ces observations démontrent les possibilités d'ingénierie en combinant des portions de différentes PKS *via* ces petits domaines d'interaction.

Des études portant sur des alignements de séquences de ces DD ont mis en avant l'existence de trois classes de DD ayant des origines phylogénétiques différentes<sup>78</sup>. Néanmoins une quatrième classe de DD existe, mais celle-ci est classée à part puisqu'elle a été mise en évidence au sein de la PKS/NRPS biosynthétisant la

tubulysine, à l'interface entre deux sous-unités type NRPS<sup>79</sup>. Les trois classes de DD présentent aux extrémités des sous-unités PKS sont constituées d'hélices  $\alpha$ , formant des structures avec des motifs en coiled-coil, essentielles à l'homodimérisation des DD et au docking des deux homodimères de DD. Les DD présents aux jonctions NRPS/NRPS (et NRPS/PKS) sont quelque peu différents et sont constitués d'hélice  $\alpha$  mais également de brins  $\beta$  qui sont essentiels à l'interaction des DD. L'interaction entre les DD est spécifique puisque l'interface est constituée de contacts électrostatiques et hydrophobes permis par des résidus clés. Ceci permet le transfert du polykétide en cours de biosynthèse de manière précise et efficace.

Les trois classes de DD des systèmes PKS ont été regroupées récemment en seulement deux classes : la classe 1 comportant les types a et b ainsi que la classe 2. Les classes 1a et 1b sont deux classes de DD de structure proche. Les DD des systèmes PKS/NRPS constituent alors la classe 3<sup>80</sup>. Les classes de DD intervenant au niveau d'interactions PKS/PKS sont décrites à la suite.

### VII.1 Les domaines de docking de classe 1a

L'étude structurale des DD a commencé avec des représentants de la classe 1a. Le DD présent en Cter comprend de 80 à 100 résidus et est composé d'une région conservée de trois hélices  $\alpha$  (hélices 1, 2 et 3), tandis que celui en Nter est composé d'une région conservée d'une hélice  $\alpha$  allant de 30 à 40 résidus (hélice 4 de la fusion). La première structure de DD en complexe a été résolue par RMN lors de l'étude des DD qui permettent l'interaction entre les deuxième et troisième sous-unités de DEBS<sup>81</sup>. Afin d'obtenir les deux DD liés entre eux, ceux-ci ont été fusionnés avec un linker de deux résidus. La structure globale se découpe en deux domaines A et B : le domaine A est constitué des hélices  $\alpha$  1 et 2 et le domaine B des hélices  $\alpha$  3 et 4 (Figure 47). Les PKS étant homodimériques, le domaine A forme donc un fagot de quatre hélices  $\alpha$ . Ce domaine est décrit comme un élément de dimérisation, ce qui permet la stabilisation de l'extrémité Cter de la deuxième sous-unité de la PKS (Figure 47). Ce domaine est séparé du domaine B par une boucle de longueur variable. Au niveau du domaine B, l'hélice  $\alpha$  4 forme un dimère de coiled-coil stabilisant l'extrémité Nter de la troisième sous-unité, à l'instar des hélices  $\alpha$  1 et 2. L'hélice  $\alpha$ 3 se fixe alors à ce coiled-coil, ce qui engendre l'interaction des deux sous-unités *via* la formation d'un fagot de quatre hélices  $\alpha$  de topologie différente de celui impliqué à la dimérisation. Cette structure

met en avant les DD comme élément d'interaction mais également de dimérisation (Figure 47).

La dimérisation au niveau du domaine A est permise grâce à des interactions hydrophobes dues à des résidus clés bien conservés dans cette classe de DD. Au niveau du domaine B, le coiled-coil est maintenu par des interactions hydrophobes grâce aussi à des résidus conservés. Ces observations montrent que les motifs de dimérisation en Cter et Nter sont conservés. A l'intérieur de ce domaine B, l'élément d'interaction est dû à des résidus hydrophobes entre les hélices  $\alpha$  3 et 4, mais également à deux ponts salins aux positions clés de l'interface qui entraînent une fidélité dans l'interaction<sup>81</sup>.

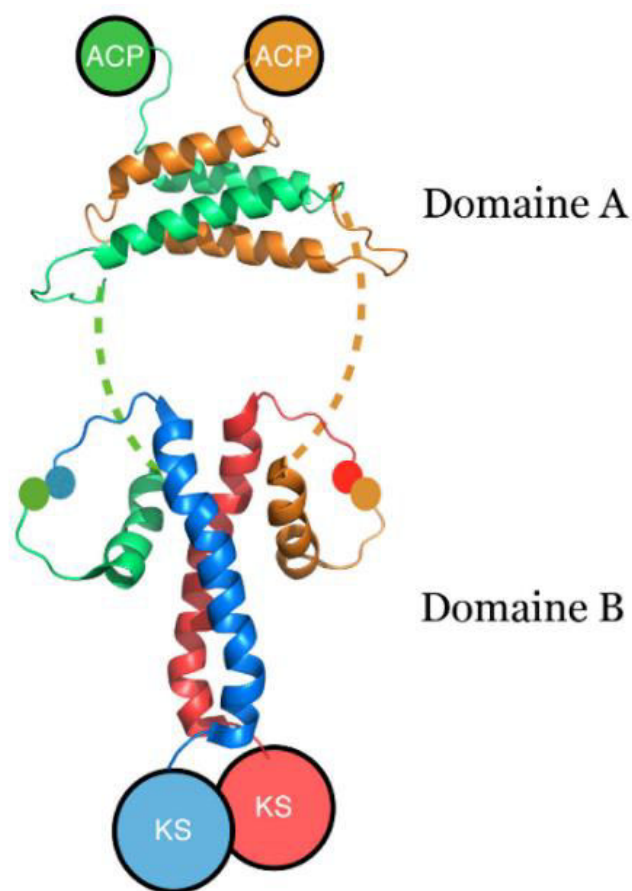


Figure 47 : Représentation de la structure des DD de l'extrémité Cter de la deuxième sous-unité et de l'extrémité Nter de la troisième sous-unité de DEBS. Le domaine A est composé des hélices  $\alpha$  1 et 2 qui se dimérisent et forment un fagot de quatre hélices  $\alpha$ . Le domaine B est, quant à lui, composé de l'hélice  $\alpha$  4 qui se dimérise pour former une structure en coiled-coil où l'hélice  $\alpha$  3 vient entrer en interaction pour former un fagot de 4 hélices  $\alpha$  de topologie distincte<sup>81</sup>.

Grâce à cette structure, de nouvelles études d'ingénierie ont été réalisées sur cette classe de DD. Au lieu d'échanger les DD en entiers, des échanges d'hélices  $\alpha$  ont été réalisés<sup>82</sup>. Ils ont été effectués au niveau des hélices  $\alpha$  3 et 4, celles directement responsables de l'interaction. Ces changements ont été pratiqués sur une PKS

minimale qui comporte le module de chargement, le module 1, le module 3 et le domaine TE de DEBS. L'ACP présent au niveau du module 1 a été remplacé par l'ACP<sub>2</sub>. Les modules 1 et 3 interagissent donc ensemble *via* les DD liés à l'ACP<sub>2</sub> et le KS<sub>3</sub>. Cette PKS minimale permet la formation d'une trikétide lactone. En remplaçant les hélices  $\alpha$  3 et 4 des DD par des hélices qui interagissent nativement, comme celles présentes entre la deuxième et la troisième sous-unité de DEBS, la formation de produit qui découle de l'efficacité d'interaction est comparable à la PKS native contenant les hélices  $\alpha$  de la première et deuxième sous-unité de DEBS. En les remplaçant par des hélices  $\alpha$  qui n'interagissent, comme celle de la première sous-unité et de la troisième, la formation du produit est plus faible ce qui signifie que le transfert du polykétide – et donc la reconnaissance des partenaires – n'est pas optimal. Ces résultats montrent que quelques changements à l'intérieur d'un couple de DD permettent de modifier la spécificité d'interaction.

A partir de cette même construction, le but a été de modifier la spécificité d'interaction non pas en échangeant les domaines DD entiers ou les hélices  $\alpha$  présentes à l'intérieur de ces domaines, et spécifiques de l'interaction, mais de faire de la mutagenèse dirigée sur les résidus impliqués dans la formation de ponts salins. Les résidus chargés R3486 et E9 respectivement présents au niveau des DD de la première et de la deuxième sous-unité de DEBS sont impliqués dans la spécificité d'interaction. L'inversion d'une des deux charges diminue ou abolie l'interaction alors que l'inversion des deux charges permet de rétablir l'interaction et ce de façon comparable au sauvage<sup>83</sup>. L'abolition de l'interaction lorsque la charge du glutamate est inversée proviendrait de la perte d'interaction entre le DD en Cter et l'ACP en amont. Ces résultats montrent qu'après la combinaison de sous-unités de PKS différentes et l'inversion de DD entiers ou seulement de portions, la mutagenèse de résidus clés de la spécificité s'avère être un bon moyen pour la production de PKS hybride.

## VII.2 Les domaines de docking de classe 1b

Les DD de classe 1b présentent une faible similarité de séquence avec les DD de classe 1a. La structure de cette deuxième classe de DD a donc été déterminée. Le choix des DD s'est porté sur ceux intervenant en Cter de la troisième sous-unité et en Nter de la quatrième sous-unité de la PKS biosynthétisant la pikromycine<sup>84</sup>. De la même

manière que les DD de la PKS de l'érythromycine A, les DD de ce système ont été fusionnés et la structure a été résolue par cristallographie aux rayons X.

Cette classe de DD est également caractérisée par la présence de quatre hélices  $\alpha$  : trois hélices en Cter et une hélice en Nter. Les hélices  $\alpha$  1 et 2 forment le domaine A alors que les hélices  $\alpha$  3 et 4 forment le domaine B. Toutefois cette fusion initiale de DD n'a pas cristallisé, à cause probablement de la flexibilité de la boucle séparant les hélices  $\alpha$  2 et 3. Ainsi la structure obtenue a été celle de la fusion des hélices  $\alpha$  3 et 4. L'hélice  $\alpha$  4 forme un dimère de coiled-coil où l'hélice  $\alpha$  3 vient se lier de la même manière que les DD de classe 1a (Figure 48). La différence dans ce système est que l'hélice  $\alpha$  3 est scindée en deux petites hélices qui entrent toutes les deux en interaction avec le dimère de coiled-coil. Il y a également une différence des résidus clés impliqués dans les interactions hydrophobes et électrostatiques. Le cœur de l'interaction entre les deux hélices  $\alpha$  est hydrophobe mais il présente également des ponts salins spécifiques à ce couple de DD.

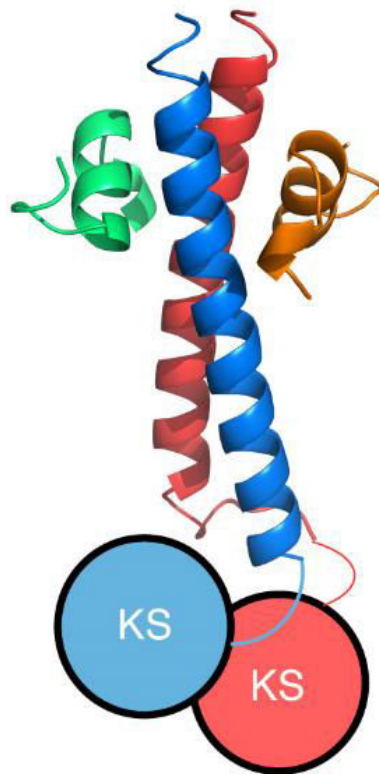


Figure 48 : Représentation de la structure du domaine B des DD comprenant l'extrémité Cter de la troisième sous-unité et l'extrémité Nter de la quatrième sous-unité de la PKS de la pikromycine. L'hélice  $\alpha$  4 (en bleu et rouge) se dimérise et forme un motif en coiled-coil où l'hélice  $\alpha$  3 (vert et orange) vient se lier<sup>84</sup>.

### VII.3 Les domaines de docking de classe 2

Les DD de classe 1a ou 1b sont présents au niveau des PKS d'actinobactéries. Des alignements de séquences de ces types de DD avec des DD issus des PKS de cyanobactéries et myxobactéries montrent une faible similarité de séquence. Ces DD dit de type 2 sont plus courts en ce qui concerne la partie en Cter, avec une taille de 40 à 60 résidus, suggérant que le domaine de dimérisation est absent, tandis que la partie en Nter est de taille identique<sup>80</sup>. Afin d'obtenir des informations sur l'interaction des DD de type 2, deux structures de DD fusionnés par un linker de 8 résidus, ont été résolues par cristallographie aux rayons X<sup>80</sup>. Il s'agit des domaines de docking qui lient les sous-unités G et H et les sous-unités K et L de la PKS biosynthétisant la curacine.

Ces deux couples de DD ont la même structure. Les extrémités Cter et Nter sont composées de deux hélices  $\alpha$  séparées par une boucle : les hélices  $\alpha$  1 et 2 pour l'extrémité Cter et les hélices  $\alpha$  3 et 4 pour l'extrémité Nter. La structure globale est dimérique et est constituée d'un fagot de 8 hélices  $\alpha$  (Figure 49). Les hélices  $\alpha$ 4 dimérisent en formant un motif en coiled-coil assurant ainsi l'homodimérisation de l'extrémité Nter. Cette structuration est possible grâce aux résidus hydrophobes présents au sein des hélices  $\alpha$ 4. Les hélices  $\alpha$ 2 se lient, quant à elles, à chacune des hélices du dimère de coiled-coil et les hélices  $\alpha$ 1 fixent l'hélice  $\alpha$ 3 ainsi que la 4. De nombreux résidus hydrophobes conservés sont impliqués dans l'interaction de ces DD tout comme les ponts salins provenant de résidus polaires non-conservés qui engendrent une interaction spécifique. La surface d'interaction résultant de ces résidus pour cette classe de DD est supérieure à celle de la classe 1 (1800 Å<sup>2</sup> contre 1100 Å<sup>2</sup>). Cependant, cette différence de superficie n'impacte pas la constante d'affinité comprise entre 2 et 20  $\mu$ M pour les deux classes de DD.

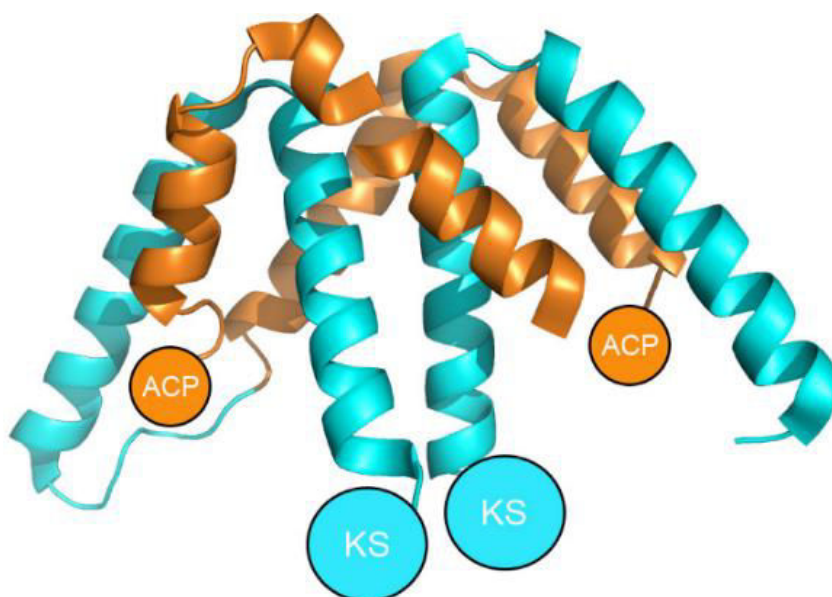


Figure 49 : Représentation de la structure des DD de classe 2 présents en Cter et Nter des sous-unités G et H de la PKS biosynthétisant la curacine. Chaque partenaire est composé de deux hélices  $\alpha$ . L'hélice  $\alpha_4$  forme un coiled-coil permettant la dimérisation de la sous-unité en Nter. L'hélice  $\alpha_2$  de l'autre partenaire se lie à ce coiled-coil tout comme l'hélice  $\alpha_1$ . L'hélice  $\alpha_1$  se lie également à l'hélice  $\alpha_3$ <sup>80</sup>.

**Un de mes objectifs de thèse a été de caractériser les domaines de docking présents à l'interface des sous-unités VirA et VirFG de la PKS biosynthétisant la virginiamycine, d'un système de PKS de type *trans*.**

## VII. Le processus de maturation post-PKS

Lors de la biosynthèse des polykétides, la PKS assure la formation du squelette du polykétide. Le polykétide doit, dans la plupart des cas, être mûri par des enzymes dites « post-PKS » afin de devenir fonctionnel. Ces enzymes ne font pas partie de la PKS bien que les gènes qui les codent se situent majoritairement à l'intérieur du même cluster que ceux codant la PKS. Par exemple, lors de la formation de l'érythromycine A, la PKS est responsable de la biosynthèse du 6-désoxyérythronolide B. Puis cinq protéines vont agir successivement sur le produit de la PKS de manière à obtenir l'érythromycine A<sup>85</sup>. Tout d'abord, l'hydroxylase EryF ajoute un groupement hydroxyle en position 6 pour donner l'érythronolide B. La glycosyltransférase, EryBV ajoute une unité mycarose, au niveau de l'hydroxyle en position 3 avec comme produit le 3- $\alpha$ -mycarosylérythronolide B. D'autres protéines dont les gènes sont à l'intérieur du même cluster conduisent à la synthèse du mycarose. Une autre glycosyltransférase, EryCIII (avec son partenaire EryCII<sup>86</sup>) ajoute ensuite une unité désosamine sur l'hydroxyle en position 5 pour former l'érythromycine D. De la même manière que pour le mycarose,

d'autres protéines dédiées vont permettre la synthèse de désosamine. Puis deux voies sont possibles : soit il y a hydroxylation en position 12 par l'hydroxylase EryK pour former l'érythromycine C puis méthylation du mycarose sur l'hydroxyle en position 3' par la méthylase EryG pour former l'érythromycine A ; ou alors il y a d'abord la méthylation par EryG puis l'hydroxylation par EryK<sup>9</sup> (Figure 50).

D'autres enzymes de modification peuvent exister mises à part les glycosyltransférases, méthyltransférases ou encore hydroxylases. On peut noter des oxydoréductases (que ce soit des oxygénases, oxydases, peroxydases, réductases ou encore déshydrogénases), des acyltransférases, des prényltransférases, des carbamoyltransférases, des aminotransférases, des halogénases ou encore des cyclases<sup>87</sup>. Au vu de l'importance de ces modifications pour l'activité du polykétide, et de la diversité des enzymes de modification, il est important de maîtriser les mécanismes de ces enzymes afin de procéder à de l'ingénierie au niveau de cette étape de biosynthèse des polykétides. Cette méthode s'insère dans les études de biologie synthétique. En effet, des échanges de gènes codant les protéines post-PKS présents au sein des clusters conduisent à la diversification des polykétides.

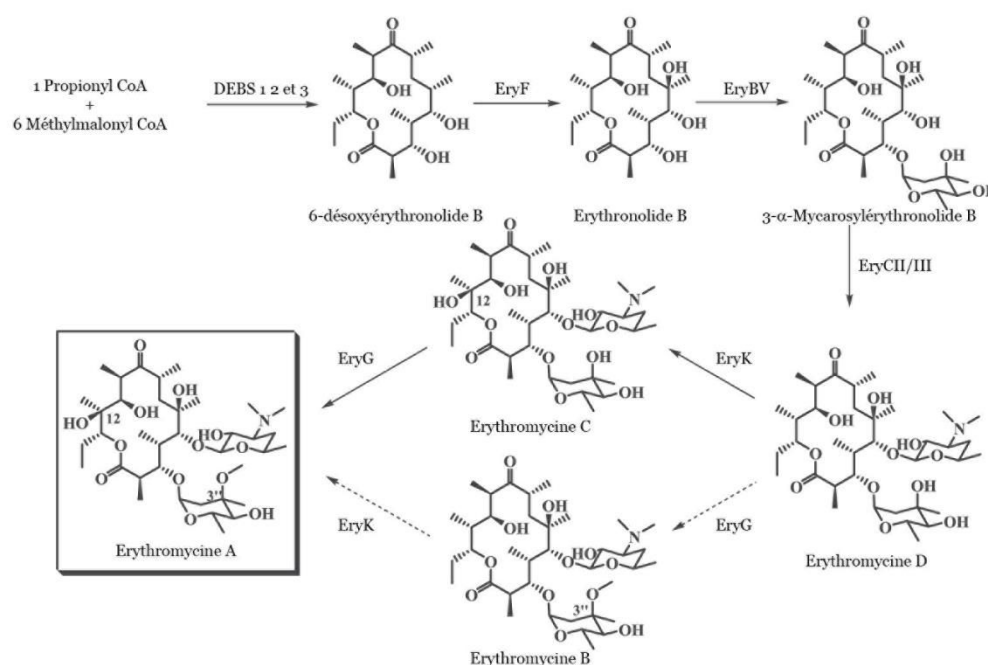


Figure 50 : Représentation du processus post-PKS de la biosynthèse de l'érythromycine A. Les enzymes EryB, EryC, EryF, EryG et EryK modifient le 6-désoxyérythronolide B, produit de la PKS, de manière à obtenir l'érythromycine A<sup>85</sup>.



Cette approche a été menée, par exemple, sur l'érythromycine A<sup>88</sup>. De nombreux pathogènes développent une résistance vis-à-vis de cette molécule. Pour passer outre cette résistance, la biosynthèse de l'érythromycine A est modifiée au niveau des deux glycosylations successives. Les voies de synthèse et de transfert de la désosamine et du mycarose sur le polykétide ont été échangées par celles du mycaminose ou de l'olivose. Ces modifications ont été réalisées chez *E. coli* et par ajout de 6-dEB dans le milieu de culture ou par ajout de 3- $\alpha$ -mycarosylérythronolide B (MEB) qui est le produit d'EryBV. Les voies de biosynthèse de mycaminose et de l'olivose ont été reconstituées chez *E. coli* et ont permis la formation de deux nouveaux dérivés de l'érythromycine A. L'un découle du MEB et est modifié par une unité mycaminose en position 5 en lieu et place de la désosamine. L'autre découle du 6-dEB et est modifié par l'olivose. Le dérivé modifié avec l'unité olivose présente une activité bactéricide contre les bactéries résistantes à l'érythromycine A. Le dérivé modifié avec l'unité mycaminose présente un profil d'action identique à celui de l'érythromycine A. Ainsi, l'utilisation des enzymes post-PKS semble être une bonne méthode pour l'ingénierie des PKS. C'est le cas des glycosyltransférases, notamment celles impliquées dans la biosynthèse de l'érythromycine A (EryBV et EryCII/III), qui présentent une large spécificité de substrat et sont capables d'accepter, dans ce cas précis, des substrats mycaminose ou olivose à la place de leurs substrats naturels désosamine et mycarose<sup>88</sup>. Cependant l'ensemble des enzymes de modification ne possède pas une spécificité large de substrat. Il est donc essentiel d'étudier les mécanismes catalytiques de ces enzymes ainsi que les résidus impliqués dans la sélection des substrats dans le but d'utiliser les enzymes post-PKS dans des études de biologie synthétique et augmenter la diversité des polykétides.

# *Objectifs*

## Objectifs

Comme nous venons de le voir, plusieurs points restent à éclaircir dans le but de réaliser des expériences de biologie synthétique des PKS de manière efficace. Au cours de ma thèse, nous nous sommes intéressés à deux d'entre eux : l'étude du contrôle de la stéréochimie ainsi que l'étude des domaines de docking.

Concernant la stéréochimie, la première partie de recherche a ciblé le changement de stéréochimie d'un polykétide modèle, *in vivo*, par modification de la PKS. Avant ces études, les travaux de Kellenberger *et al.*,<sup>68</sup> avaient consisté à insérer une KR de type épimérisante à la place de la KR non-épimérisante du module 2 de DEBS 1-TE. Il y avait formation d'un polykétide qui n'était pas réduit (signifiant que la KR n'est pas active) ou alors formation d'un produit épimérisé mais seulement sous forme de traces. Au vu de cette absence ou faible réactivité de ces deux KR, l'objectif de ma thèse était d'insérer à la place de la même KR, l'ensemble des KR de type épimérisant connu. Ceci a été fait de manière à déterminer si une épimérisation peut être introduite au niveau d'un polykétide par simple échange d'une KR. L'ensemble des PKS a été généré par génie génétique et chacune d'entre elles a été transformée dans des cellules de *S. erythraea* de manière à conduire à la formation du polykétide *in vivo*. Ces produits ont été extraits et analysés par GC-MS pour identifier leur stéréochimie.

La deuxième partie de recherche sur le stéréocontrôle a été de déterminer le mécanisme de l'épimérisation catalysée par les domaines KR. De nos jours, aucune entité n'a été mise en évidence comme étant responsable de cette réaction<sup>46</sup>. Des hypothèses suggèrent qu'une molécule d'eau positionnée à l'intérieur du site actif pourrait être responsable de la réaction ou encore la tyrosine catalytique<sup>67</sup>. Afin d'identifier l'agent responsable, nous avons voulu résoudre la structure d'un complexe ternaire d'une KR épimérisante (c'est-à-dire, en présence de son cofacteur et de son substrat (des structures des KR épimérisantes sont connues mais toujours en absence du substrat natif). Pour cela, nous avons tenté de cristalliser les KR<sub>1</sub> de la PKS pikromycine et lankamycine ainsi que la KR<sub>A</sub> de la PKS mycolactone en présence des analogues proches de leur substrat natif (dikétide-pantéthéine/phosphopantéthéine) ou plus éloigné comme *la trans*-1-décalone, ainsi qu'un dérivé réduit du cofacteur nicotinamide (NADPH<sub>3</sub>) qui est inactif, mais permet en principe d'induire le changement conformationnel associé à la fixation du cofacteur<sup>65</sup>. Nous nous sommes

également intéressés à la KR<sub>6</sub> de la PKS virginiamycine afin d'obtenir la structure d'une PKS *trans*-AT.

La troisième partie concernant le contrôle de la stéréochimie, consistait en l'identification du domaine responsable de la stabilisation du carbone en  $\beta$  du polykétide lié à l'ACP. Initialement identifié comme étant le domaine ACP<sup>69</sup>, des études plus récentes ont démontré que ce domaine n'entraîne pas en interaction avec le carbone  $\beta$  et donc ne pouvait pas stabiliser la configuration en cette position<sup>70,71,72</sup>. L'objectif était donc de confirmer que l'ACP d'un module modèle (le premier module de la PKS de la pikromycine) n'interagit pas avec cette position et de regarder si le domaine KS-AT ou le domaine KR peut être responsable de la stabilisation de la configuration. L'analyse de ces interactions a été réalisée par RMN. Les spectres HSQC de l'ACP *holo* et acylé avec le dikétide sont comparés afin d'observer l'absence de variations dans les déplacements chimiques qui confirmerait que l'ACP et son substrat n'interagissent pas. Les vitesses d'échange hydrogène/deutérium au niveau du carbone  $\alpha$  du polykétide lié à l'ACP en absence ou présence du domaine KS-AT et/ou KR sont mesurés et comparés afin d'observer une éventuelle diminution qui proviendrait de la stabilisation par un des domaines.

Enfin, concernant l'étude des domaines de docking, l'objectif a été de caractériser les DD au niveau d'un système de PKS *trans*-AT puisque l'ensemble des données et structures proviennent des PKS *cis*-AT. Le modèle choisi a été l'interface entre les modules 5 et 6 de la PKS de la virginiamycine, situés respectivement sur les sous-unités VirA et VirFG. Ici, un objectif était d'obtenir la structure par RMN des DD fusionnés entre eux et d'identifier les résidus se situant à l'interface et donc responsable de l'interaction et de sa spécificité. Un deuxième objectif était de positionner les DD par rapport au contexte modulaire et donc aux domaines voisins, par une approche SAXS, afin d'observer s'ils sont impliqués dans l'interaction.

# *Matériels et méthodes*

## Matériels et méthodes

Cette partie détaille les matériels et méthodes nécessaires au projet de cristallisation des domaines KR ainsi qu'à l'étude de la stabilisation du carbone  $\beta$  du polykétide en cours de synthèse. En ce qui concerne les matériels et méthodes utilisés lors des projets de caractérisation des domaines de docking des systèmes *trans*-AT et d'ingénierie des PKS par échange de domaines KR, ils sont décrits dans les deux publications jointes.

### I. Matériels biologiques

#### I.1 Souches bactériennes

*Escherichia coli* DH5 $\alpha$  : *supE44*,  $\Delta$ *lacU169* ( $\phi$ 80 *lacZ* $\Delta$ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1* et *relA1*. Cette souche est utilisée pour les transformations de routine. Elle permet la propagation des plasmides et possède un taux de transformation élevé.

*Escherichia coli* BL21 (DE3) : F<sup>-</sup> *ompT gal dcm lon hsdSB* (*r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>*)  $\lambda$ (DE3). Cette souche est utilisée pour l'expression et la production de protéines recombinantes. Elle contient dans son génome la séquence codant l'ARN polymérase du bactériophage T7 dont la transcription est inductible à l'IPTG *via* le promoteur *lacUV5*.

*Streptomyces venezuelae* : souche produisant la pikromycine fournie par le Professeur Joon (Ewha Womans University, Corée).

*Streptomyces virginiae* : souche produisant la virginiamycine achetée chez Genebank, Japon.

#### I.2 Vecteurs plasmidiques

Zero Blunt PCR (Invitrogen) : vecteur de clonage portant un gène de résistance à la kanamycine. Le vecteur contient le gène létal *ccdB* fusionné au gène codant la partie C-terminale de LacZ $\alpha$ . Ce vecteur permet le clonage de séquences à bouts francs.

pBG-106 : vecteur d'expression, dérivé du pET27 portant un gène de résistance à la kanamycine. Il permet la fusion d'une protéine recombinante à un tag His<sub>10</sub> en N-

terminal. Cette étiquette est clivable ensuite grâce au site de protéolyse reconnu par la protéase du rhinovirus humain 3C. La transcription de l'insert est sous le contrôle du promoteur du bactériophage T7. Il provient de l'Université de Vanderbilt.

pET20b: vecteur d'expression portant un gène de résistance à l'ampicilline. Il permet la fusion d'une protéine recombinante à un tag His<sub>6</sub> en C-terminal. La transcription de l'insert est sous le contrôle du promoteur du bactériophage T7.

pET28b : vecteur d'expression portant un gène de résistance à la kanamycine. Il permet la fusion d'une protéine recombinante à un tag His<sub>6</sub> en N-terminal. Le tag est clivable ensuite grâce au site de protéolyse reconnu par la thrombine. La transcription de l'insert est sous le contrôle du promoteur du bactériophage T7.

pET29b : vecteur d'expression portant un gène de résistance à la kanamycine. Il permet la fusion d'une protéine recombinante à un tag His<sub>6</sub> en C-terminal. La transcription de l'insert est sous le contrôle du promoteur du bactériophage T7.

pVB : vecteur d'expression portant un gène de résistance à la kanamycine. Il permet la fusion d'une protéine recombinante à un tag His<sub>6</sub> en N-terminal et au domaine B1 de la protéine G afin d'accroître la solubilité. Le tag est clivable ensuite grâce au site de protéolyse reconnu par la TEV, la protéase du virus de la gravure du tabac. La transcription de l'insert est sous le contrôle du promoteur du bactériophage T7.

### **I.3 Cosmide**

Cosmide 14F1 : partie d'ADN contenant les clusters de gènes de la PKS lankamycine et lankacidine. Il a été fourni par le Docteur Arakawa (Université d'Hiroshima, Japon),

### **I.4 Milieux de culture**

LB (Luria Bertani) : Il est composé de tryptone (20 g/L), d'extrait de levure (5 g/l), et de NaCl (5 g/L). Le pH est ajusté à 7,5 par ajout de NaOH et le milieu est autoclavé pendant 20 min à 1 bar. Pour les milieux gélosés, de l'agar est ajouté à raison de 17 g/L de milieu de culture avant stérilisation. Ce milieu est utilisé pour la croissance bactérienne d'*E. coli*.

ZYM 5052 : Il est composé de tryptone (10 g/L), d'extrait de levure (5 g/l), de Na<sub>2</sub>HPO<sub>4</sub> (25 mM), de KH<sub>2</sub>PO<sub>4</sub> (25 mM), de NH<sub>4</sub>Cl (50 mM) et de Na<sub>2</sub>SO<sub>4</sub> (5 mM). Le pH est ajusté à 7,4 par ajout de NaOH et le milieu est autoclavé pendant 20 min à 1 bar. Ensuite, sont ajoutés dans le milieu du glycérol (0,5 %), du glucose (0,05 %), du lactose (0,2 %), du MgSO<sub>4</sub> (1 mM), du FeCl<sub>3</sub> (50 µM), du CaCl<sub>2</sub> (20 µM), du MnCl<sub>2</sub> (10 µM), du ZnSO<sub>4</sub> (10 µM), du CoCl<sub>2</sub> (2 µM), du CuCl<sub>2</sub> (2 µM), du NiCl<sub>2</sub> (2 µM), du Na<sub>2</sub>MoO<sub>4</sub> (2 µM), du Na<sub>2</sub>SeO<sub>3</sub> (2 µM) et du H<sub>3</sub>BO<sub>3</sub> (2 µM). Ce milieu est utilisé pour la croissance bactérienne d'*E. coli* et est auto-inductible<sup>89</sup>.

SOC : Il est composé de tryptone (20 g/L), d'extrait de levure (5 g/L), de NaCl (5 g/L), de KCl (2,5 mM), de MgCl<sub>2</sub> (10 mM), de MgSO<sub>4</sub> (10 mM), et de glucose (20 mM). Le milieu est autoclavé pendant 20 min à 1 bar. Ce milieu est utilisé pour la régénération des bactéries compétentes après transformation.

M9 : Il est composé de Na<sub>2</sub>HPO<sub>4</sub> (6 g/L), KH<sub>2</sub>PO<sub>4</sub> (3 g/L), NaCl (0,5 g/L) et <sup>15</sup>NH<sub>4</sub>Cl (1 g/L). Le pH est ajusté à 7,4 par ajout de NaOH et le milieu est autoclavé pendant 20 min à 1 bar. Ensuite, sont ajoutés dans le milieu de la thiamine et de la riboflavine (50 mg/L), du glucose <sup>13</sup>C ou <sup>12</sup>C (4 g/L), du CaCl<sub>2</sub> (100 µM) et du MgSO<sub>4</sub> (2 mM) filtrés stérilement. Ce milieu minimum est utilisé pour le marquage isotopique des protéines.

ISP2 : Il est composé d'extrait de levure (4 g/L), d'extrait de malt (10 g/L) et de glucose (4 g/L). Le pH est ajusté à 7,3 par ajout de NaOH et le milieu est autoclavé pendant 20 min à 1 bar. Ce milieu est utilisé pour la mise en culture des souches de *Streptomyces*. Pour les milieux gélosés, de l'agar est ajouté à raison de 17 g/L de milieu de culture avant stérilisation.



## II. Clonage, production et purification des protéines recombinantes

### II.1 Extraction de l'ADN génomique de différentes espèces de *Streptomyces*

Les ADN génomiques de *S. virginiae* et *S. venezuelae* ont été extraits afin d'amplifier les séquences codant les domaines KR. Les deux souches ont été mises en culture dans du milieu ISP2 durant 48 h afin d'obtenir une quantité suffisante de cellules. L'extraction a été réalisée à l'aide du kit Wizard Genomic DNA Purification (Promega). Le protocole se décompose en cinq étapes : les cellules préalablement cultivées en milieu liquide ou solide ont été lysées. Les ARN contaminants ont alors été dégradés par l'ajout de RNase (60 µg/mL). Ensuite les protéines ont été précipitées et éliminées par centrifugation. L'ADN génomique a alors été précipité par ajout d'éthanol à 95 %, puis lavé à l'éthanol à 70 %. Une fois l'éthanol éliminé après centrifugation, les culots contenant l'ADN génomique ont été repris dans de l'eau. Ils ont alors été dosés à 260 nm et conservés à -20 °C.

### II.2 Clonage des ORF codant les protéines étudiées

Le clonage de la KR<sub>1</sub> de la PKS pikromycine a été réalisé à partir de l'ADN génomique de *S. venezuelae*, la KR<sub>1</sub> de la PKS lankamycine à partir du cosmide 14F1, et la KR<sub>6</sub> de la PKS virginiamycine à partir de l'ADN génomique de *S. virginiae*. Concernant la KR<sub>A</sub> de la PKS mycolactone, elle a été fournie, clonée dans le vecteur d'expression pVB, par le Professeur Leadlay (Université de Cambridge, UK) tout comme les protéines PanK (pantéthéine kinase) et PPAT (phosphopantéthéine adényltransférase) clonées dans le vecteur pET29b et la DPCK (3'-déphosphocoenzymeA kinase) clonée dans le vecteur pET20b utilisées dans la synthèse de coenzyme A<sup>90</sup>. Enfin les séquences codant les domaines ACP et KS-AT du premier module de la pikromycine ont été synthétisées (DNA 2.0, USA) et sous-clonées dans le vecteur pET28b.

#### II.2.1 Amplification, purification et ligation

La technique de PCR permet de réaliser l'amplification d'un fragment d'ADN compris entre deux oligodésoxyribonucléotides, dont les séquences sont complémentaires aux extrémités du fragment d'ADN à amplifier. Les

oligodésoxyribonucléotides permettent également d'introduire des sites de restriction. Les sites de restriction utilisés ici ont été : *EcoRI* pour l'oligodésoxyribonucléotide en 5' et *BamHI* pour l'oligodésoxyribonucléotide en 3' pour les KR des PKS pikromycine et lankamycine et *BamHI/HindIII* pour la KR de la PKS virginiamycine (Tableau 1).

KR <sub>1</sub> pikromycine	TTTCCGGGATCCGGCGACGACTGGCGCTAC
	TTTCCGGAATTCTCAGTCCCCTGCGTCGATGATGC
KR <sub>1</sub> lankamycine	TTTCCGGGATCCGACGCGGCCGACGCGCTG
	TTTCCGGAATTCTCATGCGCCGTGCGCCCGGCG
KR <sub>6</sub> virginiamycine	TTTCCGGAATTCCACCGAACCGTTCTGCACCG
	TTTCCGAAGCTTTCAGGGAGCGGCGGGAGC
Oligonucléotide universel 67	GTTTTCCCAGTCACGAC
Oligonucléotide universel 122	CAGGAAACAGCTATGAC

Tableau 1 : Liste des oligodésoxyribonucléotides utilisés pour ces deux projets. Les sites de restriction sont soulignés.

L'enzyme utilisée pour la réaction de PCR a été l'ADN polymérase Phusion (Finnzymes Ozyme) (Tableau 2). La présence et la taille de l'ADN amplifié ont été vérifiées par dépôt sur gel d'agarose 1 % supplémenté en GelRed (Biotium), un agent intercalant qui fluoresce aux UV. Lorsque l'ADN amplifié migre à la taille attendue, ce dernier a été extrait du gel et ensuite purifié grâce au kit Nucleobond Extract II (Macherey Nagel). Cette extraction est basée sur l'utilisation de colonnes échangeuses d'anions, l'ADN étant chargé négativement. L'ADN purifié a alors été inséré dans le plasmide Zero Blunt PCR. La formation des liaisons phosphodiester entre les extrémités de l'insert et du vecteur est catalysée par l'ADN ligase du bactériophage T4 (MBI Fermentas). La ligation s'est faite en présence d'un excès molaire d'insert de trois fois par rapport au vecteur linéarisé. Les mélanges de ligation ont été placés à température ambiante pendant une nuit.

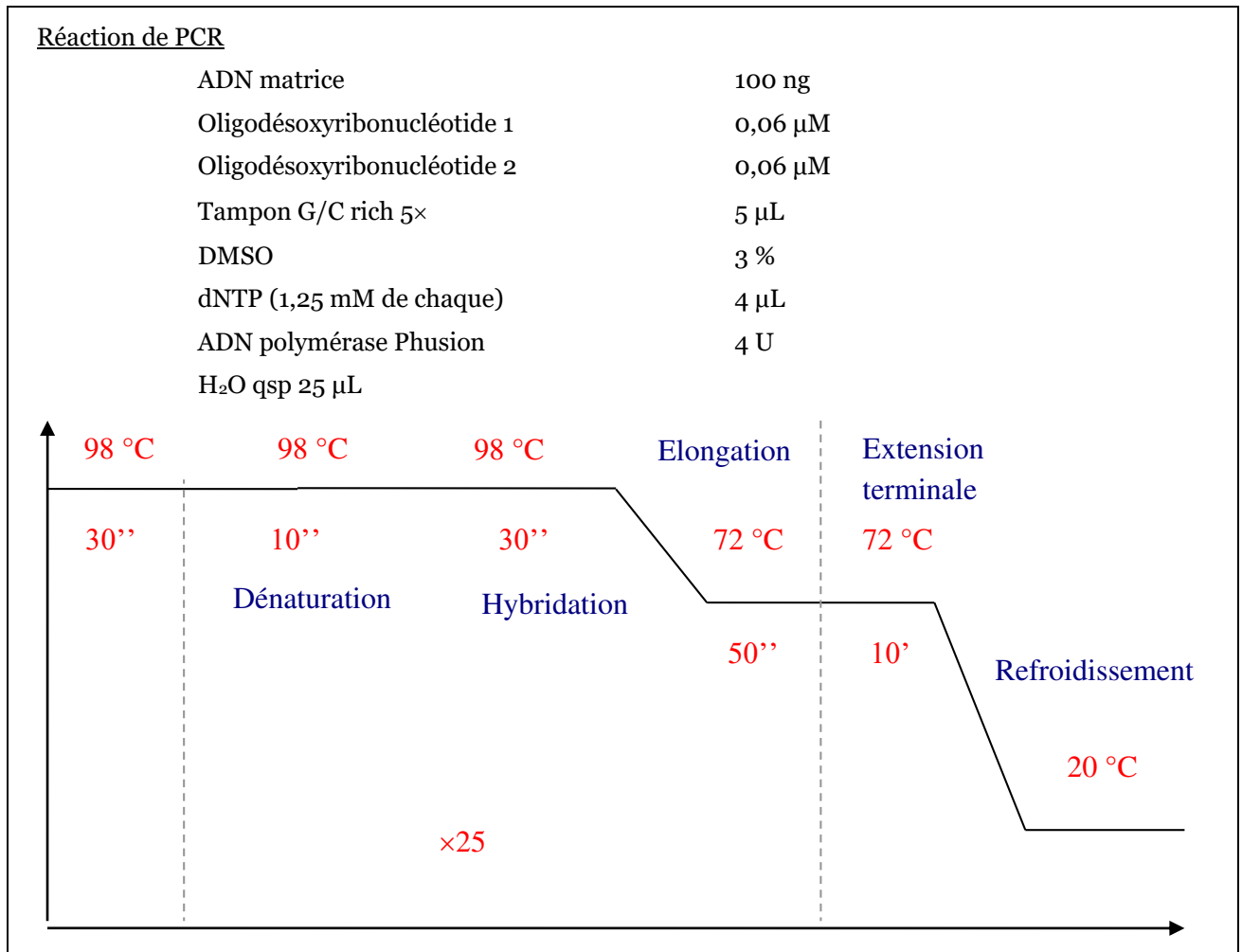


Tableau 2 : Protocole de PCR.

### II.2.2 Transformation bactérienne et sélection des vecteurs recombinants

La préparation de cellules compétentes d'*E. coli* a été faite pour pouvoir transformer la souche d'*E. coli* DH5 $\alpha$ . Une culture bactérienne a été réalisée dans 100 mL de milieu LB à 37 °C sous agitation. Cette dernière a été arrêtée en phase exponentielle de croissance lorsque la DO<sub>600nm</sub> atteint 0,5. La culture a été centrifugée et le culot repris dans une solution de TfbI (CH<sub>3</sub>COOK 30 mM, CaCl<sub>2</sub> 10 mM, KCl 100 mM, MnCl<sub>2</sub> 50 mM, et glycérol 15 % (pH 7,0)). Après une incubation de 10 min dans la glace et une nouvelle centrifugation, le culot a été repris dans 4 mL de solution TfbII (MOPS 10 mM, CaCl<sub>2</sub> 75 mM, KCl 10 mM, et glycérol 15 % (pH 7,0)). Des fractions aliquotes ont rapidement été congelées par immersion dans de l'azote liquide et stockées à -80 °C.

Les cellules d'*E. coli* rendues compétentes ont été transformées par choc thermique avec une fraction aliquote du produit de ligation. La transformation a été réalisée en mélangeant 100 ng de plasmide et 50  $\mu$ L de bactéries compétentes. Après incubation sur glace pendant 30 min, le mélange a été soumis à un choc thermique de 35 secondes à 42 °C puis placé à nouveau sur glace pendant 5 min. Ensuite, les cellules ont été régénérées pendant 1 h à 37 °C en milieu SOC avant d'être étalées sur boîte LB contenant de la kanamycine (50  $\mu$ g/mL) et incubées une nuit à 37 °C. Quelques clones obtenus ont été mis en culture et la présence de l'insert est vérifiée par PCR sur colonie en utilisant cette fois-ci l'ADN polymérase DreamTaq (MBI Fermentas) et les oligodésoxyribonucléotides universels 67 et 122 (Tableau 1).

L'intégrité des séquences insérées a été vérifiée par séquençage (Société GATC Biotech, Allemagne).

### **II.2.3 L'extraction d'ADN plasmidique**

La midipréparation d'ADN a été utilisée pour la préparation d'une grande quantité d'ADN à partir de 200 mL de cultureensemencée par 2 mL de préculture. L'ADN plasmidique recombinant a été purifié à l'aide du kit Nucleobond Xtra Midi (Macherey Nagel).

La minipréparation d'ADN a été utilisée pour la préparation rapide d'une faible quantité d'ADN dans un but analytique, à partir de 5 mL de culture. L'ADN plasmidique recombinant a été purifié à l'aide du kit Nucleospin plasmid (Macherey Nagel).

### **II.2.4 Sous-clonage des ORF codant les protéines étudiées**

Le vecteur Zero Blunt PCR recombinant contenant un insert ainsi que le vecteur pBG-106 ont été digérés pendant 2 h à 37 °C par les enzymes de restriction *EcoRI/HindIII* ou *BamHI/HindIII* à raison de cinq unités d'enzyme par  $\mu$ g d'ADN. Le vecteur linéarisé et l'insert ont été purifiés après migration sur gel d'agarose grâce au kit Nucleobond Extract II (Macherey Nagel) puis ligués ensemble sur la nuit par l'ADN ligase du bactériophage T4 à température ambiante. Une fraction aliquote du produit de ligation a été transformée dans des cellules d'*E. coli* DH5 $\alpha$  compétentes, par choc thermique. Les cellules ont été étalées sur boîte LB contenant de la kanamycine (50  $\mu$ g/mL) et incubées une nuit à 37 °C. Quelques clones obtenus ont été mis en

culture et une minipréparation d'ADN plasmidique a ensuite été réalisée après vérification de la présence de l'insert par PCR sur colonie avec les oligodésoxyribonucléotides d'amplification de l'insert. Les vecteurs ont ensuite été transformés dans des cellules d'*E. coli* BL21 (DE3).

### II.3 Expression des protéines

Une préculture bactérienne en milieu LB (ou M9 pour les cultures en milieu minimum) contenant de la kanamycine ou de l'ampicilline (100 µg/mL), suivant la résistance, a été ensemencée sur la nuit avec des cellules d'*E. coli* BL21 (DE3) contenant le plasmide d'intérêt. Ces précultures ont permis d'ensemencer les cultures bactériennes au 1/100<sup>ème</sup>. Les milieux utilisés ont été du ZYM5052 pour l'ensemble des domaines KR et du didomaine KS-AT, du LB pour l'ACP, PanK, PPAT et DPCK ou du M9 supplémenté en <sup>15</sup>N ou <sup>13</sup>C/<sup>15</sup>N afin d'obtenir de l'ACP marquée isotopiquement, le tout complémenté en kanamycine ou ampicilline (100 µg/mL). Les cultures réalisées en milieu ZYM ont été placées sous agitation 3 h à 37 °C puis 18 h à 20 °C. Les cultures réalisées en milieu LB et M9 ont été placées sous agitation à 37 °C jusqu'à une DO<sub>600nm</sub> de 0,6. L'induction de la synthèse de l'ARN polymérase T7 a alors été initiée par ajout de 0,1 mM d'IPTG. Celle-ci conduit à l'induction de la synthèse en grande quantité d'ARN messenger correspondant à la protéine recombinante. Les cultures ont ensuite été placées à 20 °C pendant 18 h. Les cellules récoltées par centrifugation à 4000g pendant 30 min à 4 °C, ont été resuspendues dans du tampon (saccharose 25 %, Tris-HCl 50 mM, EDTA 5 mM, pH 7,5) afin de les protéger de la congélation et elles ont été à nouveau centrifugées à 4000g pendant 30 min à 4 °C puis stockées à -20 °C.

## II.4 Purification des protéines

Protéine	Tampon A	Tampon B	Tampon de filtration sur gel
KR <sub>1</sub> pikromycine KR <sub>1</sub> lankamycine KR <sub>6</sub> virginiamycine KR <sub>A</sub> mycolactone	Tris 50 mM NaCl 150 mM pH 7,5	Tris 50 mM NaCl 150 mM Imidazole 1 M pH 7,5	Tris 20 mM NaCl 150 mM pH 7,5
PanK PPAT DPCK	Phosphate 50 mM NaCl 250 mM pH 7,5	Phosphate 50 mM NaCl 250 mM Imidazole 1 M pH 7,5	
KSAT pikromycine	Phosphate 50 mM NaCl 300 mM pH 8	Phosphate 50 mM NaCl 300 mM Imidazole 1 M pH 8	Phosphate 50 mM NaCl 150 mM pH 7,6
ACP pikromycine	Tris 50 mM NaCl 400 mM pH 7,5	Tris 50 mM NaCl 400 mM Imidazole 1 M pH 7,5	Phosphate 100 mM pH 6,5

Tableau 3 : Liste des tampons utilisés pour la purification des différentes protéines.

### II.4.1 Lyse cellulaire

Les culots cellulaires ont été décongelés et repris dans 30 mL de tampon A (Tableau 3) auquel ont été ajoutés 100 unités de benzonase et 6 mM de MgSO<sub>4</sub> afin de dégrader les acides nucléiques. Les cellules ont ensuite été lysées à l'aide d'un sonicateur par 3 cycles de 2 min de sonication et 1 min d'arrêt, dans la glace. Le lysat a ensuite été centrifugé pendant 45 min à 17000g afin d'éliminer les débris cellulaires.

### II.4.2 Chromatographie d'affinité Nickel-Sépharose

Le surnageant de sonication a été filtré sur un filtre de 0,45 µm de manière à éliminer les derniers débris présents (membranes, agrégats). Le surnageant a ensuite été chargé à 4 mL/min sur une colonne Nickel-Sépharose (GE Healthcare Life Sciences) préalablement équilibrée avec le tampon A, en présence de 50 mM d'imidazole à un débit de 4 mL/min, et couplée à un système FPLC (Äkta Avant ou Äkta Pure, GE Healthcare Life Sciences). La protéine d'intérêt portant plusieurs résidus histidines à son extrémité Nter (6 ou 10 suivant les protéines), s'est fixée à la colonne *via* des interactions de coordination avec le Ni<sup>2+</sup> lié à la matrice de la colonne. Avant d'injecter l'échantillon, 50 mM d'imidazole ont été ajoutés dans le surnageant de lyse afin de limiter les fixations aspécifiques des protéines cellulaires avec la phase mobile. Une fois le surnageant chargé, la colonne a été lavée avec le tampon A supplémenté

avec 50 mM d'imidazole. L'élution s'est faite à un débit de 4 mL/min par un saut à 500 mM d'imidazole, apporté par le tampon B (Tableau 3). La structure de l'imidazole est similaire à la chaîne latérale de l'histidine, par compétition elle engendre donc la libération de la protéine. Les différentes fractions d'élution ont été déposées sur gel SDS-PAGE 12,5 % et révélées au bleu de Coomassie R-250. Les fractions contenant la protéine d'intérêt ont été rassemblées.

Pour les protéines PanK, PPAT et DPCK, la purification s'est arrêtée à cette étape. Les fractions rassemblées ont été concentrées par centrifugation à 4000g avec des concentrateurs Amicon (Merck Millipore) et congelées dans l'azote liquide pour être stockées à -80 °C.

### **II.4.3 Protéolyse du tag histidine**

Pour les différents domaines KR ainsi que pour le domaine ACP et le didomaine KS-AT, les différentes fractions rassemblées ont été mises en présence de 150 µg de protéase (TEV, thrombine ou protéase du rhinovirus humain 3C (PreScission) suivant la construction) et dialysées contre 5 L de tampon A sur la nuit dans afin d'éliminer l'imidazole.

### **II.4.4 Chromatographie d'affinité Nickel-Sépharose inverse**

Après clivage du tag histidine, l'échantillon a été chargée à nouveau sur colonne Nickel-Sépharose (GE Healthcare Life Sciences) préalablement équilibrée avec du tampon A contenant 20 mM d'imidazole à un débit de 4 mL/min. La protéine étant débarrassée de son tag, elle n'entre plus en interaction avec la phase stationnaire et se retrouve dans la fraction non retenue. Cette colonne a permis d'éliminer les contaminants qui ont été élués avec la protéine d'intérêt lors de la première chromatographie d'affinité Nickel-Sépharose puisqu'ils se sont fixés à nouveau sur la colonne lors de l'étape de chromatographie d'affinité Nickel-Sépharose inverse. La protéase utilisée est taggué histidine et a ainsi été séparée de la protéine. L'élution s'est faite à un débit de 4 mL/min par un saut à 500 mM d'imidazole, apporté par le tampon B et la fraction obtenue contient les contaminants ainsi que la protéase. La pureté de la fraction non retenue sur la colonne et qui contient la protéine d'intérêt a été évaluée par dépôt sur gel SDS-PAGE 12,5 % et révélées au bleu de Coomassie R-250.

#### **II.4.5 Filtration sur gel ou chromatographie d'exclusion de taille**

Afin d'éliminer les derniers contaminants pouvant être encore présents dans la fraction protéique, une étape de filtration sur gel a été effectuée. Elle a été réalisée sur une colonne de type Superdex 75 26/60 ou alors sur Superdex 200 26/60 pour le didomaine KS-AT. Les colonnes ont été équilibrées au préalable avec le tampon de filtration sur gel (Tableau 3) et la protéine a été concentrée à 7 mL avec un concentrateur Amicon (Merck Millipore) à 4000g pour être chargée sur la colonne par l'intermédiaire d'une boucle de 10 mL. L'élution a été réalisée avec un volume colonne à un débit de 2 mL/min. Les protéines ont été séparées suivant leur taille et les différentes fractions de 2 mL ont été analysées par gel SDS-PAGE 12,5 % et révélées au bleu de Coomassie R-250. Les fractions contenant la protéine pure ont été rassemblées et concentrées par centrifugation à 4000g avec un concentrateur Amicon (Merck Millipore). La protéine a ensuite été congelée dans l'azote liquide et conservée à -80 °C.

#### **II.4.6 Dosage des protéines**

A la fin de la purification, la concentration des protéines est mesurée. Celles-ci ont été dosées à 260 et 280 nm grâce à un spectrophotomètre Nanodrop 2000C (Thermo Scientific). Cependant, le domaine ACP ne possède pas de résidus aromatiques. Il est donc impossible de mesurer sa concentration à 280 nm. La mesure d'absorbance pour cette protéine s'est donc faite par un dosage de Bradford<sup>91</sup>. Une gamme de plusieurs concentrations en BSA a été réalisée. Un mL de réactif de Bradford a été ajouté à chaque concentration en BSA, de 10 à 125 µg/mL, dans des cuves d'absorbance. La réaction a été laissée 5 min dans l'obscurité. Le réactif de Bradford réagit avec les protéines avec un déplacement d'absorption de 465 nm à 595 nm. Ainsi, pour chaque concentration, l'absorbance a été mesurée à 595 nm à l'aide d'un spectrophotomètre afin de tracer une courbe étalon. Enfin, une dilution d'ACP a été réalisée à laquelle est ajouté 1 mL de réactif de Bradford. La réaction a également été laissée dans l'obscurité pendant 5 min et l'absorbance mesurée à 595 nm. L'absorbance ainsi obtenue a été reportée à la concentration en BSA qui correspond à la concentration de la dilution en ACP.



### **III. Vers une structure des domaines KR**

#### **III.1 Analyse de l'homogénéité de la protéine**

Afin de cristalliser une protéine, il a été essentiel de travailler avec une protéine pure mais aussi homogène en solution, c'est-à-dire sans agrégation. La pureté a été estimée par l'analyse sur gel et l'homogénéité de la solution par diffusion dynamique de la lumière (DLS pour Dynamic Light Scattering) qui ont permis d'analyser l'état du soluté en solution. En effet, il est primordial d'avoir une solution de protéine monodisperse pour pouvoir aborder une étude structurale. La diffusion dynamique de la lumière permet non seulement de déterminer la composition d'un tampon dans lequel la protéine sera stable et non agrégée, mais aussi de mesurer le rayon hydrodynamique des solutés.

La protéine a été centrifugée pendant 1 h à 13000g à 4 °C afin d'éliminer les agrégats et les poussières. La cuve de l'appareil Zetasizer NanoS (Malverne) a été remplie avec 17 µL de protéine et le résultat découle d'une dizaine d'acquisition à 25 °C. L'indice de polydispersité mesuré a permis d'évaluer l'homogénéité de la protéine récemment purifiée ou décongelée.

#### **III.2 Cristallisation**

Des essais de cristallisation ont été réalisés sur la KR<sub>1</sub> de la PKS pikromycine, la KR<sub>1</sub> de la PKS lankamycine, la KR<sub>6</sub> de la PKS virginiamycine ainsi que la KR<sub>A</sub> de la PKS mycolactone afin d'obtenir leur structure. Au laboratoire, nous utilisons la méthode de diffusion en phase vapeur et la technique de la goutte assise qui a permis l'utilisation de robots de cristallisation, ainsi que la technique de la goutte suspendue.

Les premiers essais de cristallisation ont été réalisés avec des kits commerciaux afin d'obtenir les premiers cristaux mais aussi de connaître le comportement de la protéine dans différentes conditions de pH et de force ionique, par exemple. Les kits commerciaux utilisés ont été PACT premier (Molecular Dimensions), JCSG-plus (Molecular Dimensions), Morpheus (Molecular Dimensions), Wizard (Emerald BioSystems), SaltRX (Hampton research), PGA Screen (Molecular Dimensions), PEG/Ion Screen (Hampton research), JBS Screen, HR (Hampton research), Index (Hampton research) et Crystal Screen (Hampton research). Chaque kit est constitué de 96 conditions de cristallisation. Les tests ont été réalisés dans des boîtes de 96 puits de

type Greiner dans lesquelles chacun des puits est associé à trois cupules (Figure 51). Les puits ont été remplis des conditions de cristallisation à l'aide du robot Evo 150 (Tecan). Le robot Mosquito (TTP Labtech) a été utilisé pour déposer les gouttes de cristallisation dans chacune des trois cupules en mélangeant la solution de protéine à la solution de cristallisation du puits. Trois ratios différents ont été utilisés : une goutte avec un mélange de 200 nL de solution de cristallisation et 100 nL de protéine, une goutte avec 100 nL de solution de cristallisation et 200 nL de protéine et une goutte avec 100 nL de chaque. Une fois les boîtes réalisées, elles ont été placées à 20 °C puis observées tous les jours à la lampe binoculaire pour suivre l'apparition de cristaux.

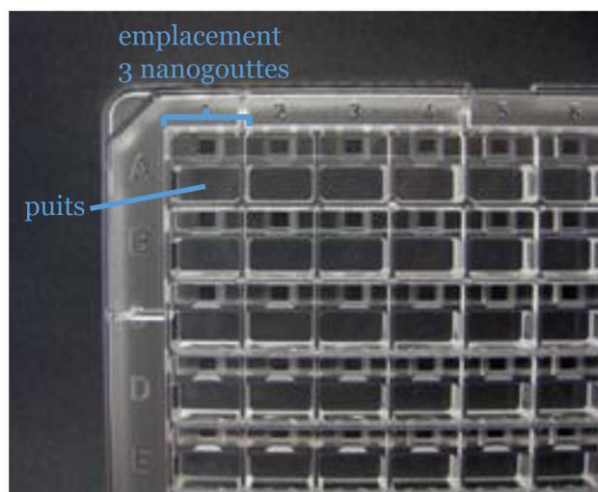


Figure 51 : Boîte utilisée lors des essais de cristallisation. Les 96 puits sont remplis avec les différentes conditions de cristallisation. Au-dessus de chaque puits, les 3 emplacements reçoivent les 3 gouttes de cristallisation avec des ratios différents (Hampton research).

Lorsqu'une protéine cristallise dans une des conditions d'un kit de cristallisation, les gouttes sont reproduites manuellement dans des boîtes Linbro de 24 puits, permettant d'augmenter leur volume de l'ordre du  $\mu\text{L}$  (Figure 52). Elles permettent d'affiner les cristaux obtenus en faisant varier les différents paramètres influant la cristallisation que ce soit la concentration en protéine, le pH, la concentration et nature de l'agent précipitant, la concentration et nature des sels, les additifs, ou le ratio solution de cristallisation/protéine. Les puits sont remplis de la solution de cristallisation et les gouttes sont déposées sur des lamelles en verre. Les lamelles sont ensuite déposées sur le puits et le tout est fermé hermétiquement par un joint en silicone. Les gouttes se retrouvent suspendues vers le puits dans un système clos qui induit des échanges entre la goutte et le puits.



Figure 52 : Représentation d'une boîte Linbro permettant l'affinement des cristaux obtenus avec les kits de cristallisation (Hampton research).

### III.3 Congélation des cristaux obtenus

Pour enregistrer des données de diffraction, il est indispensable de congeler les cristaux afin de les protéger de la détérioration due aux radicaux engendrés par la photolyse de l'eau sous le faisceau de rayon X. La congélation limite ainsi la diffusion des radicaux à l'intérieur du cristal et de ce fait limite sa dégradation. Les cristaux de macromolécules biologiques contiennent entre 30 et 70% d'eau. Il est donc indispensable de protéger les cristaux contre la formation de glace lors de la congélation, sous peine de détériorer le cristal et ainsi perdre le pouvoir de diffraction. Les cristaux obtenus ont été trempés dans la solution de cristallisation contenant 30 % de glycérol, qui joue le rôle de cryoprotectant. Ils ont été ensuite congelés dans un flux d'azote et conservés dans de l'azote liquide.

### III.4 Enregistrement et traitement des données

Les données de diffraction ont été enregistrées au synchrotron SOLEIL (Gif-sur-Yvette) sur les lignes Proxima 1 et Proxima 2. Les clichés de diffraction ont été indexés et intégrés avec le programme XDS<sup>92</sup>, puis les données mises à l'échelle par le logiciel XSCALE<sup>92</sup>. Pour l'étape de phasage par remplacement moléculaire, plusieurs programmes ont été utilisés : Phaser<sup>93</sup>, MolRep<sup>94</sup> ou encore Balbes<sup>95</sup> en utilisant comme modèles les structures les plus proches à savoir : la KR<sub>1</sub> de DEBS (fichier PDB : 2FRO), la KR<sub>1</sub> de la PKS tylosine (2Z5L), la KR<sub>2</sub> de la PKS amphotéricine (3MJE) et la KR<sub>3</sub> de la PKS pikromycine (3QP9). Cette méthode consiste à positionner la protéine dans le cristal étudié en déterminant la matrice de rotation et le vecteur de translation,

et permet de résoudre le problème de la phase en utilisant la structure d'une protéine homologue.

### III.5 Obtention de dérivés d'atomes lourds de cristaux de protéine

Des cristaux ont été produits en boîte Linbro en présence de lanthanides pour résoudre le problème de la phase, étant donné que la KR<sub>1</sub> de la PKS lankamycine ne contient que deux méthionines. Il est donc impossible d'utiliser le signal anomal du sélénium apporté par les sélénométhionines qui remplacent les méthionines de la protéine. Les lanthanides, qui regroupent les éléments avec un nombre atomique de 57 à 61, ont alors été utilisés dans les conditions de cristallisation à des concentrations allant de 10 à 100mM. Dans ce cadre, nous avons utilisé différents lanthanides comme l'euprium, l'ytterbium et le gadolinium (Eu, Yb, Gd) chélatés avec plusieurs composés : DOTA (acide 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetra-acétique), DO3A (acide 1,4,7,10-tetraazacyclododecane-1,4,7-triacétique), (DPA)<sub>3</sub> (tris(dipicolinate)) et HPDO3A (acide 10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacétique)<sup>96,97</sup> (NatXRay) (Figure 53). Les lanthanides sont des diffuseurs anomaux et permettent de résoudre le problème de la phase par SAD (Dispersion Anomale à une Seule longueur d'onde) ou MAD (Dispersion Anomale à Multiple longueur d'onde).

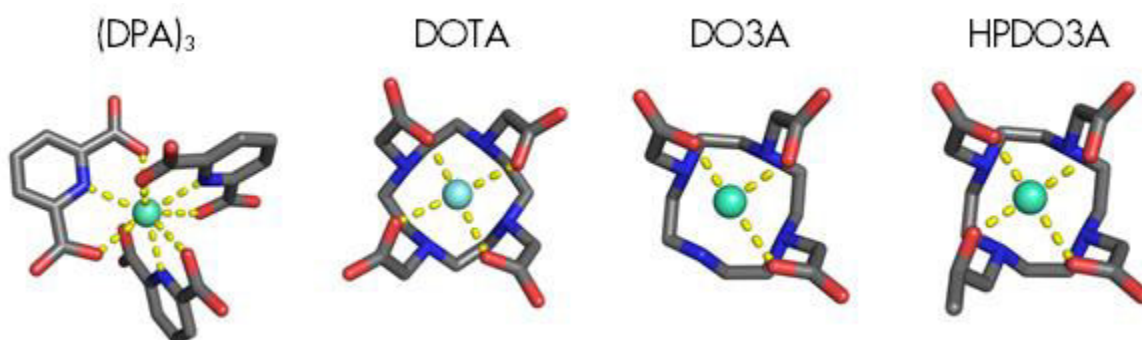


Figure 53 : Les lanthanides sont en vert et chélatés par quatre molécules différentes.

Afin d'obtenir des dérivés d'atomes lourds, deux méthodes principales existent ; le trempage et la co-cristallisation. Le trempage consiste à faire tremper les cristaux dans un mélange contenant la solution de cristallisation et des lanthanides. Les lanthanides diffusent alors à l'intérieur du cristal. La co-cristallisation consiste à faire cristalliser la protéine avec un lanthanide dans la solution de cristallisation. Cette

méthode est plutôt utilisée lorsque le trempage ne permet pas d'obtenir des dérivés d'atomes lourds car la co-cristallisation présente des défauts, notamment d'inhiber la cristallisation.

## IV. Etude des interactions entre le substrat CoA-dikétide et les domaines ACP, KS-AT et KR du premier module de la PKS pikromycine

### IV.1 Synthèse du CoA-dikétide

Les protéines PanK, PPAT et DPCK permettent la synthèse du CoA-dikétide à partir du pantéthéine-dikétide<sup>90</sup>. La protéine PanK ajoute un phosphate sur la pantéthéine-dikétide en présence d'ATP pour former la phosphopantéthéine-dikétide. La protéine PPAT ajoute de l'adénosine pour donner du déphospho-CoA-dikétide en présence d'ATP. Enfin la protéine DPCK ajoute un phosphate en présence également d'ATP avec comme produit le CoA-dikétide (Figure 54).

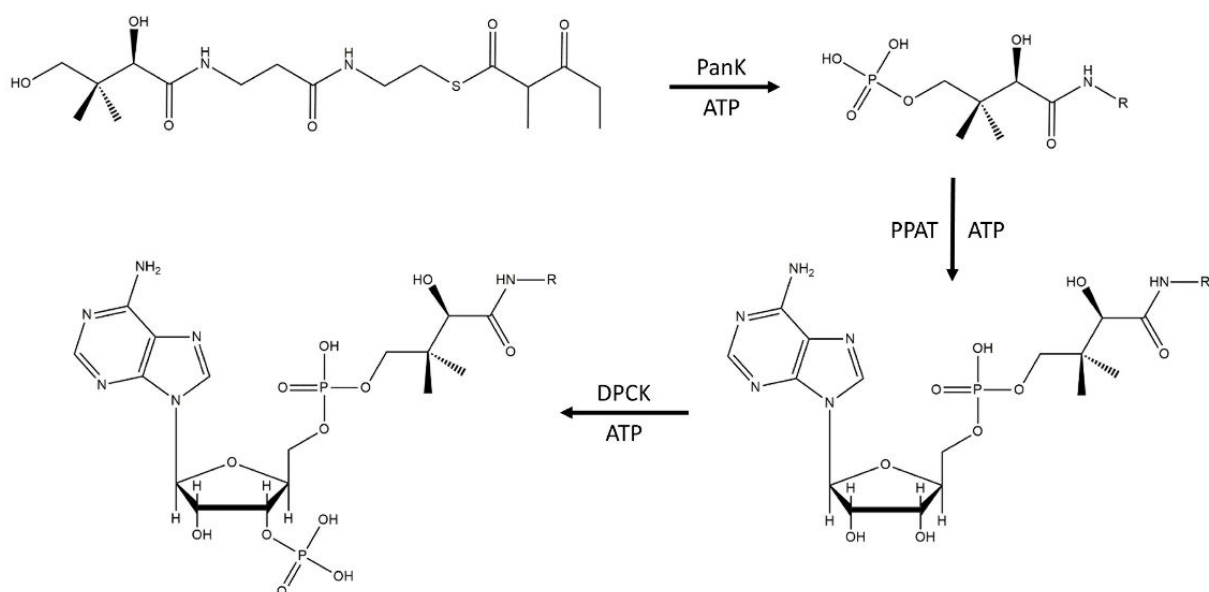


Figure 54 : Réaction de synthèse du CoA dikétide.

Afin de synthétiser du CoA-dikétide, 10 mg de dikétide pantéthéinylé (marqué isotopiquement <sup>13</sup>C ou non, synthétisé par l'équipe du Dr. A Kirschning, Université de Hannover) ont été mélangés avec 75 mg d'ATP, 75  $\mu$ M de PanK, 50  $\mu$ M de DPCK et 150  $\mu$ M de PPAT dans un volume final de 10 mL de tampon Tris-HCl 50 mM, KCl 20 mM,

MgCl<sub>2</sub> 10 mM, pH 7,5. Le mélange a été laissé à 30 °C pendant 3 h puis congelé afin d'être lyophilisé sur la nuit. L'obtention du produit final, le CoA-dikétide, a été vérifiée par analyse en spectrométrie de masse couplée à de la chromatographie en phase liquide. Ces analyses ont été réalisées en collaboration avec C. Paris (LiBio, Université de Lorraine). Le CoA-dikétide est repris dans un mélange eau 95 %, acétonitrile 5% à une concentration de 13 mM.

## IV.2 Modification de l'ACP

Dans un premier temps, la modification de l'ACP de la PKS pikromycine a été faite par du CoA. Elle a été réalisée en utilisant deux phosphopantéthéinyl transférases (PPTases) différentes : la Sfp (Surfactine phosphopantéthéinyl-transférase) de *Bacillus subtilis* (le vecteur a été fourni par le Dr B. Broadhurst, Université de Cambridge) et la Svp (*Streptomyces verticillus* PPTase) de *Streptomyces verticillus* (le vecteur a été fourni par le Dr B. Shen, Scripps Florida). A 100 µM d'ACP ont été ajoutés 3 µM de Sfp ou Svp, 5 mM de CoA, 10 mM de MgCl<sub>2</sub> et 5 mM de DTT. Le volume réactionnel a été ajusté à 300 µL avec du tampon phosphate 100 mM, pH 7,5. Les ions Mg<sup>2+</sup> sont le cofacteur des deux PPTases et ont donc été ajoutés en excès. Le DTT réduit la fonction thiol du CoA, et empêche ainsi la formation de pont disulfure entre deux molécules de CoA. La réaction a été laissée à température ambiante sur la nuit. Après la modification, il a été nécessaire de purifier l'ACP, dorénavant sous forme *holo*, pour effectuer des études par RMN. La purification réalisée sur une colonne Superdex 75 10/300 GL a permis d'éliminer la PPTase et le CoA en excès et de placer la protéine modifiée dans le tampon phosphate 100 mM pH 7,5. La présence ou non de modification a été vérifiée par spectrométrie de masse ESI-Q-TOF en collaboration avec le Dr H. Mazon au service commun de spectrométrie de masse et des techniques couplées de l'Université de Lorraine. Le spectromètre est un microTOF-Q (Bruker Daltonics). Les échantillons ont été injectés à une vitesse de 6 µL/min en même temps qu'un étalon externe, « Tuning Mix ». Les données ont été traitées par le logiciel Bruker Data Analysis (Bruker Daltonics).

Des études de séquençage par empreinte peptidique massique ont été également réalisées afin d'identifier le résidu sur lequel la modification s'est effectuée. Elles ont été faites en collaboration avec le Dr J. B. Vincourt par MALDI-TOF sur la plateforme de protéomique de la Fédération de Recherche 3209 (Bioingénierie Moléculaire,

Cellulaire et Thérapeutique, Université de Lorraine) avec un spectromètre Autoflex speed. La matrice utilisée est de l'acide  $\alpha$ -cyano-hydroxycinnamique (HCCA). Pour l'ensemble des expériences de spectrométrie de masse, les échantillons ont été préalablement dialysés par concentration dans du tampon acétate d'ammonium 50 mM, pH 6,4.

L'ACP a également été modifié par les CoA-dikétide et CoA-dikétide  $^{13}\text{C}$  synthétisés. Ces modifications ont été réalisées en utilisant uniquement la Sfp. A 180  $\mu\text{M}$  d'ACP sont ajoutés 3  $\mu\text{M}$  de Sfp, 10 mM de  $\text{MgCl}_2$ , 5 mM de DTT, 1 mM de CoA-dikétide et le volume est complété à 300  $\mu\text{L}$  avec du tampon phosphate 100 mM, pH 7,5. La réaction a été laissée à température ambiante sur la nuit. Après modification, l'ACP a été purifié par filtration sur gel sur une colonne Superdex 75 10/300 GL. Une fois l'ACP élué, il a été analysé par spectrométrie de masse couplé à de la chromatographie en phase liquide et par ionisation ESI (en collaboration avec C. Paris, LiBio, Université de Lorraine).

### IV.3 Acquisition des spectres RMN

Plusieurs spectres RMN ont été enregistrés par le Dr B. Chagot sur un spectromètre Bruker 600 MHz de la plateforme SCBIM (Service Commun de Biophysicochimie des Interactions Moléculaires, Université de Lorraine). Pour ce faire, un échantillon protéique de 100  $\mu\text{M}$  a été placé dans un tube RMN de 3 mm de diamètre auquel a été ajouté 10 % de  $\text{D}_2\text{O}$  afin de calibrer le champ magnétique du spectromètre.

Des spectres RMN<sup>98</sup> ont été enregistrés sur l'ACP1 *apo* de la PKS pikromycine afin d'attribuer les déplacements chimiques des atomes de la chaîne principale. Pour cela, la protéine a été marquée isotopiquement au  $^{13}\text{C}$  et  $^{15}\text{N}$ . Les expériences ont été enregistrées sur un échantillon concentré à 1 mM. Les spectres enregistrés ont été un spectre 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC (Figure 55a) permettant d'observer les atomes d'hydrogène et d'azote de la fonction amide du squelette peptidique et les spectres 3D HNCACB, HNCA, HNCO, HNHA et CBCAcoNH permettant d'observer les atomes de la fonction amine d'un acide aminé (i) ainsi que les carbones  $\alpha$  et  $\beta$  des acides aminés (i) et (i-1) pour le spectre HNCACB (Figure 55b) ; les atomes de la fonction amine d'un acide aminé (i) ainsi que les carbones  $\alpha$  des acides aminés (i) et (i-1) pour le spectre HNCA (Figure 55c) ; les atomes de la fonction amine d'un acide aminé (i) ainsi que le carbone

carbonyle de l'acide aminé (i-1) pour le spectre HNCO (Figure 55d) ; les atomes de la fonction amine d'un acide aminé (i) ainsi que les carbones  $\alpha$  et  $\beta$  des acides aminés (i-1) pour le spectre CBCAcoNH (Figure 55e) et les atomes de la fonction amine d'un acide aminé (i) ainsi que l'hydrogène  $\alpha$  de l'acide aminé (i) pour le spectre HNHA. Les spectres ont été analysés avec le logiciel SPARKY (T. D. Goddard et D. G. Kneller, SPARKY 3, University of California, San Francisco).

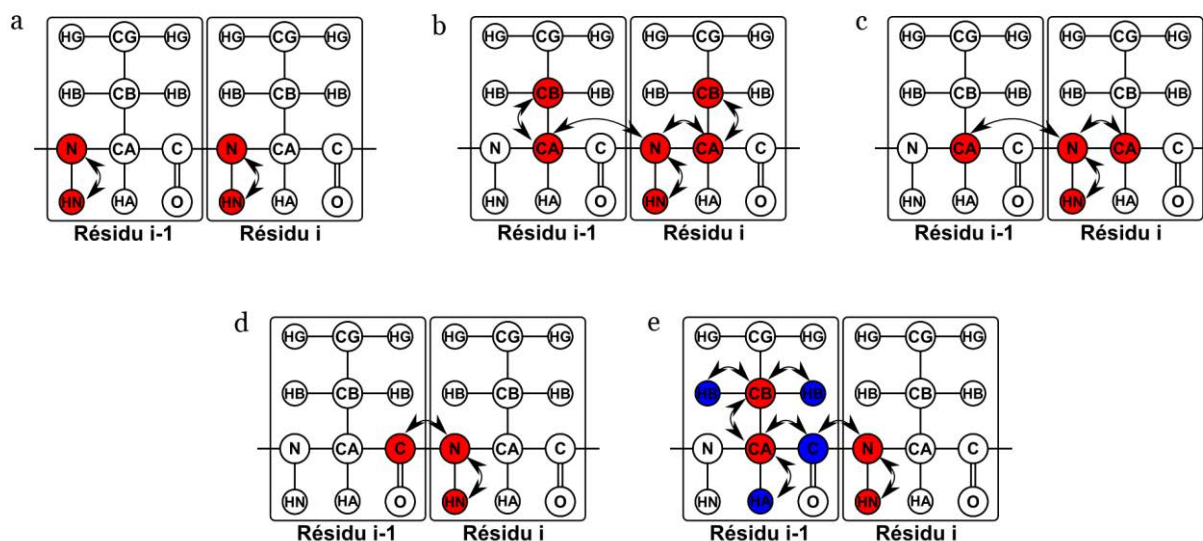


Figure 55 : Schéma récapitulatif des spectres enregistrés a)  $^1\text{H}$ - $^{15}\text{N}$  HSQC b) HNCACB c) HNCA d) HNCO e) CBCAcoNH.

Des spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC ont également été enregistrés sur un échantillon protéique marqué  $^{15}\text{N}$  concentré à  $180\ \mu\text{M}$  et modifiée avec du CoA et du CoA-dikétide. Ces spectres ont été enregistrés afin d'observer de potentielles variations de déplacements chimiques résultants de l'interaction entre le substrat et l'ACP par comparaison avec les spectres  $^1\text{H}$ - $^{15}\text{N}$  HSQC des formes *apo* de l'ACP.

Des spectres 1D  $^{13}\text{C}$  et COCA (permettant l'observation de deux carbones covalents liés respectivement à un hydrogène et un oxygène (Figure 56)) ont été enregistrés sur le CoA-dikétide  $^{13}\text{C}$  seul ainsi que sur l'ACP  $^{12}\text{C}$   $^{14}\text{N}$  modifié avec ce substrat. Le but a été de comparer les déplacements chimiques des carbones  $^{13}\text{C}$  du dikétide afin d'observer des différences de déplacements chimiques entre les atomes du substrat libre et du substrat lié à l'ACP afin de montrer une éventuelle interaction entre l'ACP et le substrat.



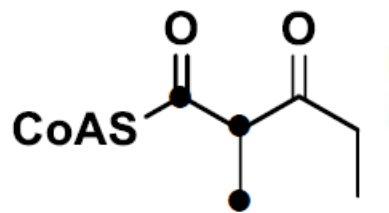


Figure 56 : Représentation du CoA-dikétide  $^{13}\text{C}$ . Les carbones marqués sont indiqués par des points noirs et leurs déplacements chimiques sont observés par les spectres  $^1\text{D}^{13}\text{C}$  et COCA.

# *Résultats*

## Résultats

### Partie 1 : Caractérisation d'une nouvelle classe de domaine de docking de PKS trans-AT

Plusieurs structures de domaines de docking (DD) ont été résolues à ce jour<sup>80,81,84</sup>. L'ensemble de ces domaines de docking provient de systèmes *cis*-AT.

Récemment, notre équipe a résolu la structure en solution par SAXS du module 5 *apo* de la sous-unité VirA de la PKS virginiamycine<sup>29</sup>. Contrairement aux alignements de séquence qui laissaient penser que l'extrémité Cter du module est un domaine de dimérisation, il s'avère que le module 5 présente une structure ouverte accessible aux partenaires. Les deux extrémités Cter n'interagissent pas, et donc les deux extrémités sont plutôt des DD.

Pour vérifier cette hypothèse, nous avons résolu la structure par RMN multi-dimensionnelle du couple de DD (le Cter de VirA (VirA<sup>CDD</sup>) et le Nter de VirFG (VirFG<sup>NDD</sup>)). Les deux DD ont été clonés et exprimés sous forme de protéine de fusion liés ensemble par un linker G<sub>3</sub>SG<sub>3</sub>S). La structure a révélé que chaque domaine de docking est composé de deux courtes hélices  $\alpha$  qui interagissent deux à deux pour former un fagot de quatre hélices  $\alpha$  de topologie unique dans la PDB.

Les études par RMN ont également montrées que le VirA<sup>CDD</sup> seul est en équilibre entre dimères non-spécifiques, alors qu'en présence d'un excès de VirFG<sup>NDD</sup> il forme un complexe stable avec son partenaire. Nous avons aussi démontré que VirFG<sup>NDD</sup> seul présente très peu de structures secondaires par dichroïsme circulaire (CD), mais il acquiert des structures secondaires supplémentaires en présence de 2,2,2-trifluoroéthanol, qui mime un environnement hydrophobe. De plus nous avons démontré par SAXS que VirFG<sup>NDD</sup> n'est pas une protéine globulaire. Cependant, il est structuré lorsqu'il est en complexe avec son partenaire VirA<sup>CDD</sup>. Toutes ces données sont cohérentes avec l'idée que VirFG<sup>NDD</sup> est une protéine intrinsèquement désordonnée (IDP), mais nous ne pouvons pas exclure qu'elle adopte une structure stable en présence de la KS en aval.

L'analyse de la structure par RMN a également permis de déterminer la nature des interactions entre VirFG<sup>NDD</sup> et VirA<sup>CDD</sup> qui impliquent des interactions hydrophobes et polaires. Le rôle des résidus a été vérifié par des études de mutagenèse dirigée couplé à la titration calorimétrique isotherme (ITC) et le CD. Nous nous sommes concentrés uniquement sur des mutants de VirA<sup>CDD</sup>, puisque le VirFG<sup>NDD</sup> se comporte comme un IDP et par conséquent, les mutations peuvent avoir une influence sur la structuration induite de ce domaine et/ou son interaction avec son partenaire. Des mutations au niveau de trois résidus du VirA<sup>CDD</sup> situées à l'interface du complexe (D6940, N6949 et E6954), ont bien diminuées son affinité pour son partenaire VirFG<sup>NDD</sup> ( $K_d$  plus élevé), soutenant la validité du modèle de docking développé sur la base de la structure du complexe covalent.

Des études par SAXS ont également été menées, afin d'identifier les positions des domaines flanquants les DD : l'ACP<sub>5b</sub> de VirA et la KS<sub>6</sub> de VirFG. La structure de VirA ACP<sub>5b</sub>-CDD et VirFG<sup>NDD</sup> fusionnés (ACP<sub>5b</sub>-CDD-VirFG<sup>NDD</sup>) montre que le domaine ACP<sub>5b</sub> n'est pas en interaction avec le complexe des DD. La fusion ACP<sub>5b</sub>-CDD-VirFG<sup>NDD</sup>-KS<sub>6</sub> permet de placer l'ACP<sub>5b</sub> à l'extrémité de la structure et met en évidence que le linker situé entre VirFG<sup>NDD</sup> et la KS<sub>6</sub> est compacte.

## Novel docking interactions mediate intersubunit communication in the virginiamycin *trans*-acyl transferase polyketide synthase

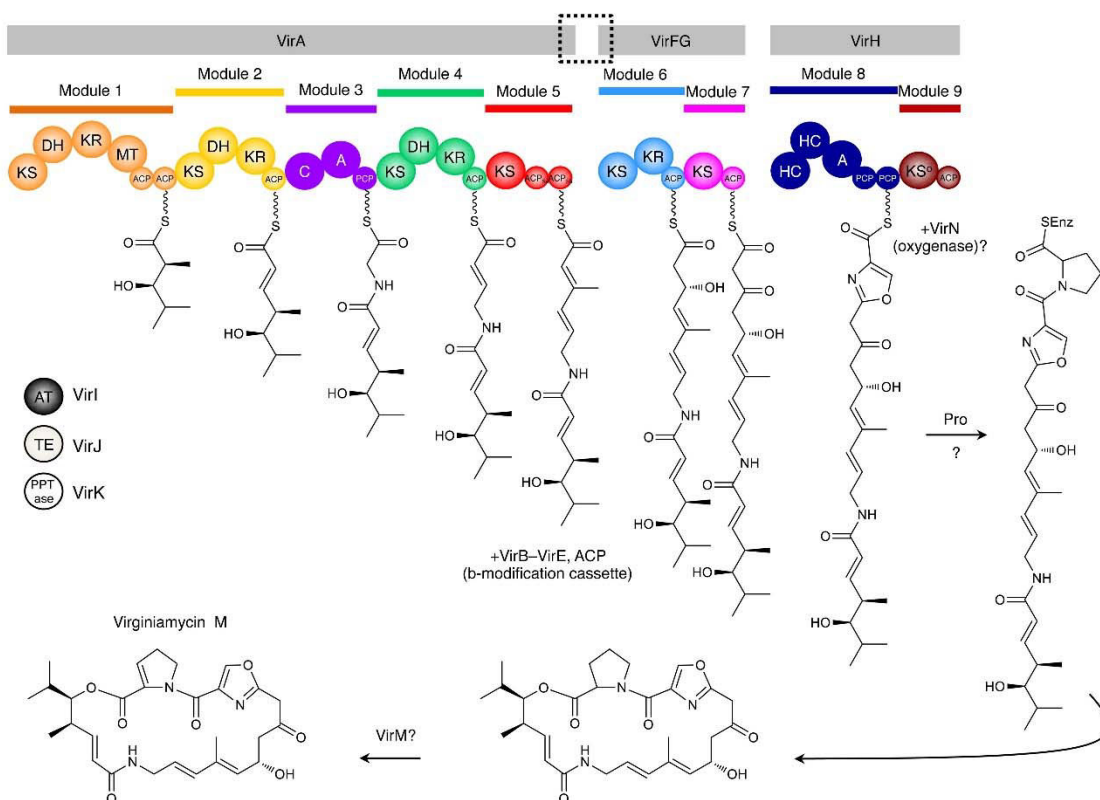
Jonathan Dorival<sup>1,4</sup>, Thibault Annaval<sup>1,4</sup>, Fanny Risser<sup>1</sup>, Sabrina Collin<sup>1</sup>, Pierre Roblin<sup>2,3</sup>, Christophe Jacob<sup>1</sup>, Arnaud Gruez<sup>1\*</sup>, Benjamin Chagot<sup>1\*</sup> and Kira J. Weissman<sup>1\*</sup>

<sup>1</sup>UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandœuvre-lès-Nancy Cedex, France. <sup>2</sup>SOLEIL Synchrotron, 91192 Gif-sur-Yvette, France. <sup>3</sup>INRA-URBIA, 44316 Nantes, France. <sup>4</sup>These authors contributed equally to this work

\*e-mail: arnaud.gruez@univ-lorraine.fr, benjamin.chagot@univ-lorraine.fr or kira.weissman@univ-lorraine.fr

**Modular polyketide synthases (PKSs) direct the biosynthesis of clinically-valuable secondary metabolites in bacteria. The fidelity of chain growth depends on specific recognition between successive subunits in each assembly line — interactions mediated by C- and N-terminal ‘docking domains’ (DDs). We have identified a new family of DDs in *trans*-acyl transferase PKSs, exemplified by a matched pair from the virginiamycin system. We show using NMR that the C-terminal DD (VirA<sup>C</sup>DD) induces folding of the N-terminal DD (VirFG<sup>N</sup>DD), the first example to our knowledge of an intrinsically disordered protein in bacterial secondary metabolism. Fusion of the two termini results in a stable fold for VirFG<sup>N</sup>DD, and an overall protein-protein complex of unique topology whose structure we support by site-directed mutagenesis. Furthermore, using small-angle X-ray scattering, the positions of the flanking acyl carrier protein and ketosynthase domains have been identified, allowing modeling of the complete intersubunit interface.**

Synthesis of structurally complex, reduced polyketides in bacteria occurs on gigantic multienzymes called polyketide synthases (PKSs)<sup>1</sup>. These systems are likened to molecular-scale assembly lines, as each round of chain extension and chemical tailoring of newly-incorporated building blocks is carried out by a specific module of enzymatic domains. Nature has evolved two distinct classes of modular PKSs – *cis*-acyl transferase (AT) and *trans*-AT. Within the majority of *cis*-AT PKSs, such as the prototypical 6-deoxyerythronolide B synthase (DEBS) responsible for assembling the core of erythromycin A, all of the component domains are present within the gigantic polypeptides<sup>2</sup>. *Trans*-AT PKSs, on the other hand, are characterized by one or more free-standing enzyme activities (including notably the ATs) which collaborate *in trans* with the PKS multienzymes to accomplish both chain building and processing (**Fig. 1**)<sup>3,4</sup>. They also characteristically include modules of non-ribosomal peptide synthetase (NRPS) which incorporate amino acids into the growing intermediates.



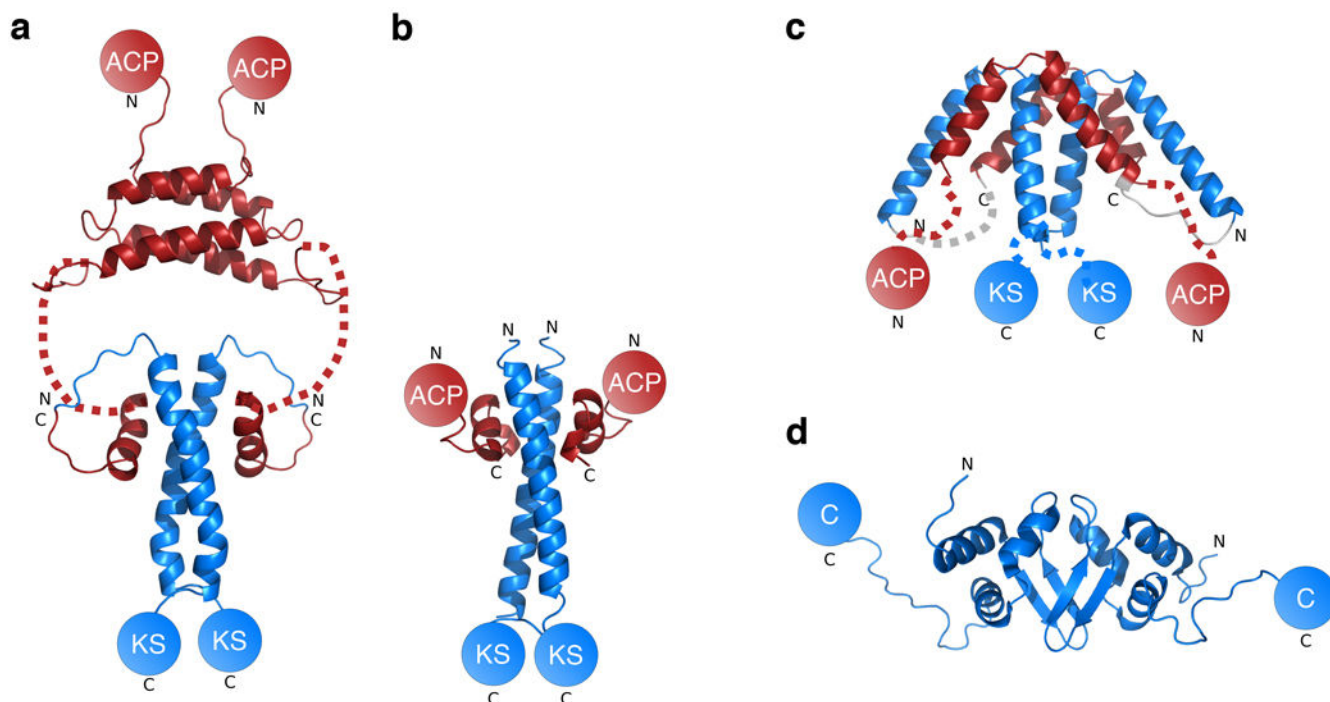
**Figure 1 | The virginiamycin (Vir) *trans*-AT polyketide synthase.** The PKS comprises three subunits, VirA, VirFG (whose N-terminal sequence was recently revised by us<sup>21</sup>) and VirH, as well as additional modules which were not identified in the initial cluster description<sup>4</sup> for starter unit selection and introduction of proline. The ? indicate additional steps where the enzyme responsible is not yet clear. The system incorporates many features characteristic of *trans*-AT PKS systems, including a *trans*-acting acyl transferase VirI, duplicated domains (acyl carrier proteins (ACPs) of modules 1 and 5, peptidyl carrier proteins (PCPs) of module 8), non-ribosomal peptide synthetase modules (3 and 8), inactive domains (KS<sup>o</sup> of module 9), and a cassette of enzymes which together introduces a  $\beta$ -methyl functionality. The experiments described here reveal the molecular details of the interface between subunits VirA and VirFG (boxed).

One conserved feature of *cis*- and *trans*-AT PKSs is the distribution of the functional modules among multiple polypeptide subunits<sup>5</sup>. Thus, generation of native polyketide products critically depends on the ability of the subunits to arrange themselves in the correct order. Conversely, rational manipulation of interpolypeptide interfaces to change the sequence of the subunits represents an attractive strategy for generating polyketide analogues, as no modifications are made to the component modules. Indeed, these are increasingly viewed as discrete operational units, given the high interdependence of the functional domains<sup>6,7</sup>.

In *cis*-AT PKS, it has been shown that communication at intersubunit junctions is mediated by portable recognition sequences called ‘docking domains (DDs)’, located at the extreme termini of the subunits, and these have already been exploited to generate polyketide analogues<sup>8–12</sup>. Three evolutionarily-related docking domain classes have been identified to date in PKS systems<sup>13</sup>, as well as a fourth class which operates at NRPS–NRPS/–PKS junctions in mixed PKS–NRPSs<sup>14</sup>. The NMR or X-ray crystal structures of covalently-fused complexes of the PKS docking domains have been solved<sup>15–17</sup> (**Fig. 2**), and the N-terminal DDs shown to adopt the same structures when expressed together with the downstream ketosynthase domains<sup>17–19</sup>. The close structural homology between the two actinobacterial classes of DDs prompted their grouping into ‘Class 1’ (these will be referred to here as 1a and 1b), while the third class, which is present in myxobacteria and cyanobacteria, has been designated as ‘Class 2’<sup>17</sup> (by extension, the mixed PKS–NRPS DDs constitute ‘Class 3’).

Class 1 and 2 DDs share two central features: i) the N-terminal docking domains are homodimeric and form a coiled-coil motif which serves as a platform for docking of the likewise  $\alpha$ -helical C-terminal docking domains; and ii) the docking interfaces are largely hydrophobic in character and complementary in shape, while specific charge:charge interactions at critical positions contribute to docking specificity. One notable difference between the two classes is that Class 1 C-terminal docking domains incorporate two  $\alpha$ -helices upstream of the docking  $\alpha$ -helix, which together form a dimerization motif (**Fig. 2a**). Class 2 C-terminal docking domains, on the other hand, comprise only two  $\alpha$ -helices, both of which associate with the N-terminal docking domain. Furthermore, the coiled-coil of the Class 2 N-terminal DD is preceded by an additional  $\alpha$ -helix, and thus docking gives rise to an overall eight  $\alpha$ -helical bundle (**Fig. 2c**), in contrast to the four  $\alpha$ -helical bundles characteristic of Class 1 DDs. Despite this architectural divergence and an approximately 2-fold difference in interface size, the affinity of association for both DD classes when the immediate upstream/downstream functional domains are present is remarkably consistent, with  $K_d$ s in the 2–20  $\mu$ M range<sup>16,17,20</sup>.

We recently characterized by small angle X-ray scattering (SAXS)<sup>21</sup> the structure of *apo* module 5 from the VirA subunit of the virginiamycin hybrid *trans*-AT PKS-NRPS<sup>4</sup> of *Streptomyces virginiae* (**Fig. 1**). In the course of this work, we noted that the extreme C-terminal sequence (designated 'VirA<sup>CDD</sup>') showed homology to a region from the CurA subunit of the curacin *cis*-AT PKS, which had been described as a 'dimerization motif'<sup>22</sup>. However, comparison by SAXS of *apo* module 5 in the presence and absence of this region did not reveal a change in the structure of the C-terminus, as the module retained its open form. This prompted us to investigate whether VirA<sup>CDD</sup> was rather a docking domain, mediating a specific interaction with a partner DD located on subunit VirFG ('VirFG<sup>NDD</sup>'). We show here that VirA<sup>CDD</sup> and VirFG<sup>NDD</sup> constitute the founding members of a new family of docking domains from *trans*-AT PKSs. Using a multidisciplinary approach combining NMR, site-directed mutagenesis, circular dichroism (CD), isothermal titration calorimetry (ITC) and SAXS, we have characterized their interaction in detail. We have furthermore employed SAXS to locate the immediate upstream acyl carrier protein (ACP) and downstream ketosynthase (KS) domains in positions that are fully compatible with our recent SAXS analysis of virginiamycin module 5 (ref. 21). Furthermore, modeling of the overall VirA/VirFG interface strongly suggests that interaction between successive subunits establishes a closed reaction chamber for chain extension, a fundamentally different mode of operation to *cis*-AT PKSs.



**Figure 2 | Three classes of docking domains from *cis*-AT PKSs.** (a) NMR structure of a complex of covalently-fused Class 1a docking domains (PDB: 1PZQ, 1PZR)<sup>15</sup>. The C-terminal DDs (red) comprise three  $\alpha$ -helices, the first two of which form a four  $\alpha$ -helical bundle which acts as a dimerization element. The third helix wraps around the coiled-coil formed by the partner N-terminal docking domain (blue). The long, flexible linker joining the second and third  $\alpha$ -helices of the C-terminal DD is represented as a dashed line. C and N indicate C- and N-termini, respectively. (b) X-ray crystal structure of a covalent complex of Class 1b docking domains (PDB: 3F5H)<sup>16</sup>. Although by sequence analysis, this class also incorporates a C-terminal dimerization motif, it was not included in the investigated construct. The overall structure resembles that in a, but the C-terminal DD comprises two smaller  $\alpha$ -helices and the precise hydrophobic and polar interactions at the interface differ. (c) X-ray crystal structure of a covalent complex of class 2 docking domains (PDB: 4MYI)<sup>17</sup>. The C-terminal DD does not incorporate a dimerization region. Instead, its two  $\alpha$ -helices associate with two  $\alpha$ -helices provided by the N-terminal docking domain to form an overall 8  $\alpha$ -helical bundle. (d) NMR structure of a class 3 N-terminal docking domain (PDB: 2JUG)<sup>14</sup>. Each protomer within the DD homodimer comprises three  $\alpha$ -helices and two  $\beta$ -strands, which adopt an overall  $\alpha\beta\beta\alpha$  topology. For all docking domain complexes, charge:charge interactions are critical determinants of specificity. Key to the domains up- and downstream of the docking domain complexes: ACP, acyl carrier protein; KS, ketosynthase; C, condensation.

## RESULTS

### *In silico* analysis of docking in *trans*-AT PKSs

In earlier work, we identified VirA<sup>C</sup>DD as comprising the 50 residues C-terminal of the conserved ACP<sub>5b</sub> domain of module 5 (Fig. 1)<sup>21</sup>. Similarly, inspection of the N-terminus of VirFG revealed 70 amino acids N-terminal to the conserved KS domain as a candidate docking domain partner (Supplementary Results, Supplementary Fig. 1a). Although BLAST<sup>23</sup> analysis of this region returned no homologues within the PDB, the first 28 residues of VirFG<sup>N</sup>DD show convincing sequence homology to VirA<sup>C</sup>DD and to multiple Class 2 docking domains<sup>17</sup> (Supplementary Fig. 1b,c). The C-terminal 42 amino acids, on the other hand, are not conserved, suggesting no involvement in docking. These observations prompted us to consider whether homologous sequences were present at intersubunit junctions in additional *trans*-AT PKS, allowing <sup>C</sup>DD and <sup>N</sup>DD families to be identified in these systems for the first time.

In *cis*-AT PKSs, interpolypeptide junctions are formed between partner docking domains with additional contacts contributed by the flanking C-terminal ACP and N-terminal KS domains<sup>6,15-17,24</sup>. Our analysis of 39 published *trans*-AT PKSs reveals that such interfaces are formed by a much greater diversity of domain pairs (Supplementary Table 1). It is also striking that in the majority of cases, the domains comprising the junction are in fact members of a single chain extension



module (i.e. the functional module is split between two different multienzyme polypeptides). Nonetheless, in a number of cases, the intersubunit interfaces are formed by the ends of modules. For a subset of such interfaces, analysis of the sequences directly downstream from the conserved ends of the ACP domains, and those directly upstream from the conserved start sites of the N-terminal domains (KS, ER or condensation (C)/heterocyclization (HC)), revealed regions with low but suggestive homology to VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD, respectively. All of the regions were predicted by secondary structure analysis (PsiPred<sup>25</sup>) to contain two  $\alpha$ -helices or one long  $\alpha$ -helix (**Supplementary Fig. 2**).

The identified junctions (**Supplementary Table 2**) include the ACP/KS interfaces present in the kirromycin, oxazolomycin (2 junctions), chivosazol, thailandamide and enacyloxin (4) systems, and the ACP/C or ACP/HC (PKS/NRPS) interfaces within the kirromycin, oxazolomycin, chivosazol, thailandamide (which incorporates an analogous ACP–ACP/KS interface), rhizopodin and disorazol assembly lines. *Streptomyces pristinaespiralis* also harbors a hybrid *trans*-AT PKS-NRPS which generates virginiamycin (though the product is referred to as pristinamycin IIA), and as expected, sequences with homology to VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD were present on the respective subunits SnaE2 and SnaE3<sup>26</sup>; nonetheless, the <sup>N</sup>DD equivalent of SnaE3 was N-terminally truncated, suggesting incorrect identification of the *snaE3* site start. Indeed, inspection of the genomic region of *S. pristinaespiralis* (GenBank: FR682001) revealed a putative alternative start codon 147 base pairs 5' of the published start site (new start site at base pair 65054); translation of this longer region gave a sequence incorporating a complete N-terminal docking domain (**Supplementary Fig. 2b**). Conversely, the published N-terminus of TaiM from the thailandamide PKS<sup>27</sup> begins some 300 residues upstream of a Met and the first conserved  $\alpha$ -helix of the putative docking domain, again suggesting an incorrectly assigned start site. On the other hand, the TaiE subunit from the same PKS also contains a long extension N-terminal to the docking domain, but this cannot apparently be explained by a mis-assigned start codon.

Our analysis also revealed two putative additional families of docking domains at ACP/X interfaces which do not show strong homology to the VirA<sup>C</sup>DD/VirFG<sup>N</sup>DD sequences (**Supplementary Table 2**), a set of junctions which might incorporate docking domains but for which no convincing homologues exist within the other published clusters, and multiple interfaces which apparently contain no docking domains.

### VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD serve as docking domains

Having identified VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD as representative of their respective families of C- and N-terminal docking domains, we aimed to confirm their interaction. For this, we expressed and purified (**Supplementary Fig. 3**) the following constructs in recombinant form from *E. coli* BL21(DE3) as His<sub>6</sub>-SUMO-tagged proteins using plasmid pBG102 (Vanderbilt University Center for Structural Biology): VirA<sup>C</sup>DD, VirFG<sup>N</sup>DD, VirA ACP<sub>5b</sub>-<sup>C</sup>DD, VirFG<sup>N</sup>DD-linker (all 70 residues up to the conserved KS<sub>6</sub>), VirFG<sup>N</sup>DD-KS<sub>6</sub>, as well as a protein (VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD) in which the two docking domains were covalently fused via the same 8-amino acid linker (G3SG3S) used previously to study the Class 2 docking domains of *cis*-AT PKSs<sup>17</sup> (**Supplementary Methods**, and for the sequences of all constructs see **Supplementary Table 3**). In order to allow efficient monitoring of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD by UV-Vis during purification, we added a tyrosine residue at the N- and

C-terminal ends, respectively – a position which we judged unlikely to interfere with the native docking interaction involving the opposite termini. Following cleavage of the His<sub>6</sub>-SUMO-tag using human rhinovirus 3C (H3C) protease, the N-termini of all of the proteins also harbored a non-native GPGS sequence. In the case of VirFG<sup>NDD</sup>-KS<sub>6</sub>, we designed the expression construct to include 179 residues downstream of the KS<sub>6</sub> domain (encompassing consecutive regions which show homology to the so-called KS-AT and post-AT linkers observed in structures of KS-AT didomains from *cis*-AT PKSs<sup>18</sup>), as this choice of boundary allowed us to obtain the KS of module 5 as a stable protein<sup>21</sup>.

As an initial demonstration that the docking domains interact, we carried out analytical gel filtration in the presence of VirA<sup>CDD</sup>, VirFG<sup>NDD</sup> and a 1:1 mixture of the two domains (all at an initial concentration of 670 μM). On their own, VirA<sup>CDD</sup> and VirFG<sup>NDD</sup> gave elution volumes of ca. 13.7 and 14.5 ml, respectively, while the 1:1 mixture gave a single peak with an elution volume of 14 ml (**Supplementary Fig. 4**), providing the first evidence for a specific interaction between the two docking domains. Conversion into molecular weights yielded the following data: VirA<sup>CDD</sup>, 11.8 kDa (monomer weight 4.3 kDa); VirFG<sup>NDD</sup> 7.4 kDa (monomer weight 4.1 kDa) and the VirA<sup>CDD</sup>/VirFG<sup>NDD</sup> complex, 9.9 kDa (heterodimeric complex weight, 9.9 kDa). These data suggested that both domains are oligomeric when isolated, but recombine to form a 1:1 complex.

Measurement of the affinity of the interaction by isothermal titration calorimetry (ITC) yielded a value of  $5.8 \pm 0.2 \mu\text{M}$  (from three independent analyses) (**Supplementary Fig. 5a**), which is within the range measured for Class 1 and Class 2 docking domains (4–20 μM)<sup>16,17</sup>; a closely-similar  $K_d$  (4 μM) was measured for the VirA<sup>ACP<sub>5b</sub>-CDD</sup>/VirFG<sup>NDD</sup> pair (**Supplementary Fig. 5b**). Inclusion of the 42-residue linker downstream of VirFG<sup>NDD</sup> did not significantly alter the affinity ( $K_d = 5.5 \mu\text{M}$  for the interaction between VirA<sup>CDD</sup> and VirFG<sup>NDD</sup>-linker (**Supplementary Fig. 5c**)), supporting our supposition that this region is not involved in docking.

### Characterization of docking between VirA<sup>CDD</sup> and VirFG<sup>NDD</sup> by NMR, SAXS and allied methods

To investigate in detail the interaction between the two docking domain partners, we analyzed <sup>15</sup>N-labeled VirA<sup>CDD</sup> in isolation and in the presence of unlabeled (<sup>14</sup>N) VirFG<sup>NDD</sup>, so that in either case, only signals arising from VirA<sup>CDD</sup> would be visible. In the absence of its partner docking domain, the [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum of VirA<sup>CDD</sup> (100 μM) exhibited only 15 of 37 expected peaks (**Supplementary Fig. 6a**). In contrast, when a 2–4-fold excess of <sup>14</sup>N VirFG<sup>NDD</sup> was also present, 36 of 37 peaks were observed (**Supplementary Fig. 6b**). Two explanations account for the fewer than expected number of peaks in the spectrum of unpartnered VirA<sup>CDD</sup>: i) a monomer-dimer equilibrium for self-association of VirA<sup>CDD</sup>, as observed previously<sup>21</sup>, occurring on an intermediate NMR time-scale; and/or, ii) an equilibrium between folded and unfolded states, also on the intermediate time-scale. We favor explanation i, as data obtained by ITC as the concentration of VirA<sup>CDD</sup> was increased incrementally are consistent with a net monomer to dimer transition, with the protein dimerizing completely at 30 μM (**Supplementary Fig. 5d**). To investigate whether VirA<sup>CDD</sup> retains its structure even in the monomeric form, we analyzed it by CD at concentrations both above (100 μM) and below (10 μM) the dimerization

point. This experiment showed that while auto-association did increase the overall secondary structure content, VirA<sup>C</sup>DD retained 73% of its  $\alpha$ -helical content in monomeric form (**Supplementary Fig. 7a**).

Although isolated VirA<sup>C</sup>DD clearly homodimerizes, in the presence of excess VirFG<sup>N</sup>DD it preferentially forms a stable, docked complex, explaining the observation of almost all anticipated NMR peaks. This fact allowed us to fully assign <sup>15</sup>N, <sup>13</sup>C-VirA within the context of this complex using a combination of 2D and 3D NMR experiments, in order to determine its secondary structure (**Supplementary Fig. 6b**). Consistent with the secondary structure predictions, bound VirA<sup>C</sup>DD comprises two short  $\alpha$ -helices (residues Q6936–R6947 and I6953–L6960; numbering based on GenBank: BAF50727)).

We next carried out these experiments in reverse, with <sup>15</sup>N-labeled VirFG<sup>N</sup>DD and unlabeled (<sup>14</sup>N) VirA<sup>C</sup>DD. The [<sup>1</sup>H, <sup>15</sup>N]-HSQC spectrum of isolated VirFG<sup>N</sup>DD (100  $\mu$ M) showed the expected number of peaks, but with poor chemical shift dispersion (**Supplementary Fig. 6c**). Dynamical studies also showed it to be highly flexible (**Supplementary Fig. 6d–f**). Taken together, these results indicated that VirFG<sup>N</sup>DD alone lacks a well-defined 3D structure, and so could be an intrinsically disordered protein (IDP)<sup>28</sup>. (For a discussion of the applicability of these findings to other *trans*-AT PKS, see **Supplementary Note**). In agreement with this, CD analysis yielded data consistent with only partial structuration (22%  $\alpha$ -helix; **Supplementary Fig. 7b**), while analysis by ITC under the same conditions as for VirA<sup>C</sup>DD provided no evidence for dimerization (**Supplementary Fig. 5e**). As further support for the IDP character of VirFG<sup>N</sup>DD, addition of 2,2,2-trifluoroethanol (TFE) to the buffer significantly increased the  $\alpha$ -helical content as measured by CD (**Supplementary Fig. 7c**). This reagent mimics the hydrophobic environment experienced by proteins in protein-protein interactions, and so is widely exploited to reveal IDPs with a propensity to undergo induced folding<sup>29</sup>. We also analyzed VirFG<sup>N</sup>DD by SAXS under the NMR conditions, yielding an  $R_g$  of  $16 \pm 1$  Å and a  $D_{max}$  of 66 (**Supplementary Fig. 8a** and **Supplementary Table 4**). The  $D_{max}$  value is consistent with an extended, albeit partially-folded protein (predicted  $D_{max}$  for a fully-extended protein of 40 residues = 160 Å), as is the average 3D molecular form determined using DAMMIN<sup>30</sup> ( $\chi^2 = 2.32$ ) (**Supplementary Fig. 8b**). The form of a Kratky plot<sup>31</sup> of the obtained data also confirmed that the protein is not globular – i.e. that it does not fold stably – but showed that it is nonetheless likely to contain residual  $\alpha$ -helix (**Supplementary Fig. 8c**).

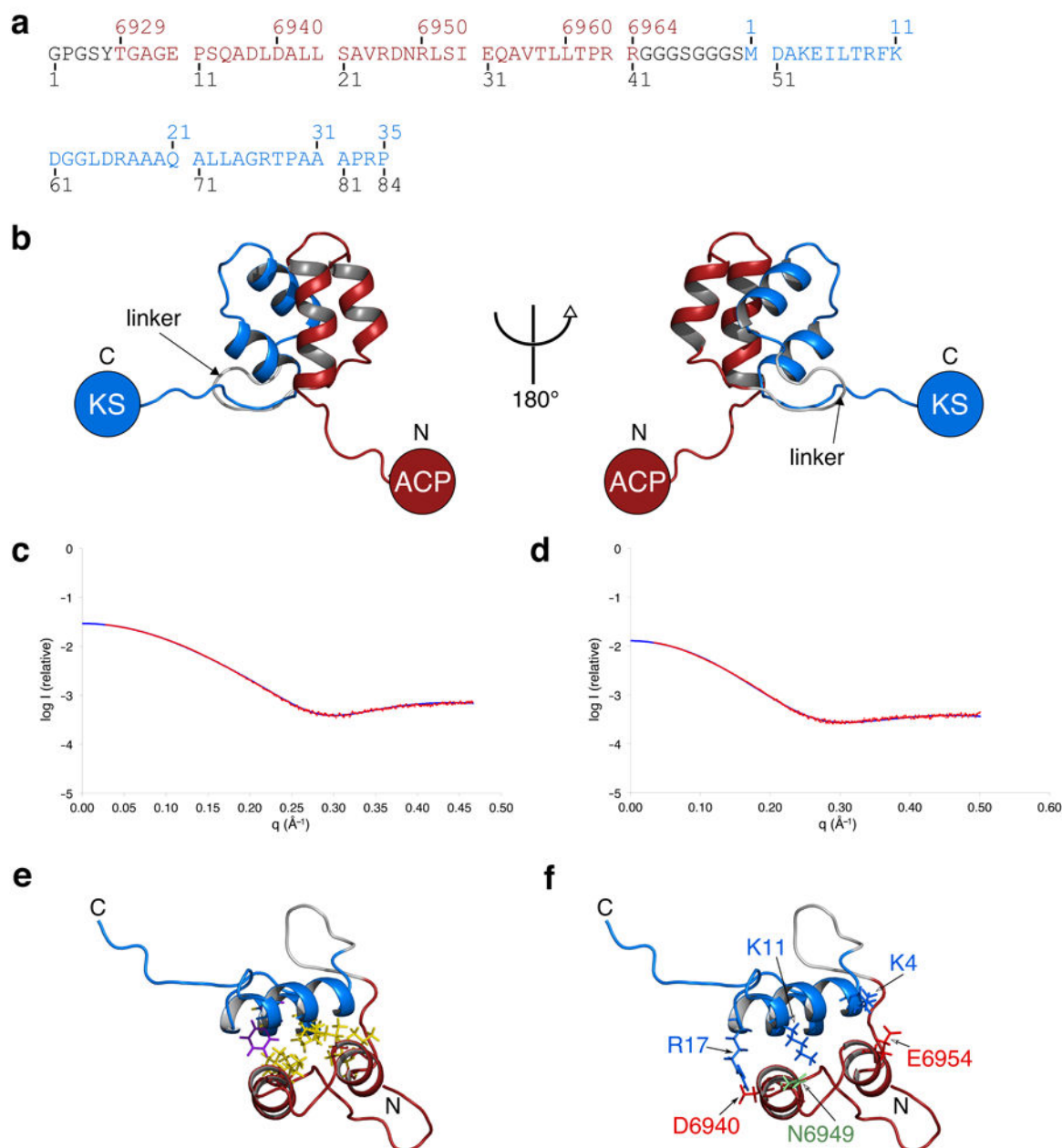
Addition of a four-fold excess of <sup>14</sup>N-VirA<sup>C</sup>DD (ca. 400  $\mu$ M) provoked a number of spectral modifications, including significant changes in chemical shift and an overall increase in chemical shift dispersion (**Supplementary Fig. 6g**). These types of spectral changes are entirely consistent with a coupled process of binding/induced structuration of VirFG<sup>N</sup>DD by VirA<sup>C</sup>DD. To bolster the case that VirFG<sup>N</sup>DD was gaining structure upon contact with VirA<sup>C</sup>DD, we carried out differential CD analysis, acquiring spectra of VirFG<sup>N</sup>DD and VirA<sup>C</sup>DD alone and in complex (**Supplementary Fig. 7d**). This analysis showed that the spectrum of the complex does not correspond to a simple average of the spectra of the two individual partners ( $\alpha$ -helical content increased from 40% to 52% (CDNN software<sup>32</sup>)). VirA<sup>C</sup>DD is already well-structured at the concentration used in this experiment due to self-association, and thus the secondary structure increase seen when the partners are combined likely results from induced folding of VirFG<sup>N</sup>DD. An essentially identical [<sup>1</sup>H, <sup>15</sup>N]-HSQC spectrum of <sup>15</sup>N-labeled VirFG<sup>N</sup>DD was obtained in the presence of VirA ACP<sub>5b</sub>-<sup>C</sup>DD (**Supplementary Fig. 6h**), consistent with the result from ITC showing that the presence of the ACP does not impact on the docking domain interaction.

We followed up these observations by elucidating the complete structure of the docking domain complex, using the construct VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD in which the two docking elements were covalently fused<sup>17</sup> (for the sequence of this protein, see **Fig. 3a**). Structural studies of <sup>15</sup>N,<sup>13</sup>C-labeled VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD (**Fig. 3, Supplementary Fig. 6i,j**) were carried out by multi-dimensional heteronuclear NMR spectroscopy. Structure calculation was performed using a two-step procedure consisting of initial structure generation using CYANA<sup>33</sup> followed by restrained molecular dynamics refinement within Amber<sup>34,35</sup>. The quality of the structures was assessed using PROCHECK-NMR<sup>36</sup> and the MolProbity<sup>37</sup> server (**Supplementary Table 5**). The first important result is that the introduced linker between the two docking domains is unstructured in the complex (**Fig. 3b**). Consistent with this, amide chemical shifts arising from this region are clustered together in the [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum, and dynamical studies show it to be highly mobile (**Supplementary Fig. 6k**). Taken together, these data confirm that the linker acts simply as a flexible tether, and does not contribute to the docking interface, arguing that our fusion strategy has faithfully captured the native interaction.

Within the context of this fusion construct, VirA<sup>C</sup>DD comprises two consecutive, short  $\alpha$ -helices as found for discrete VirA<sup>C</sup>DD in complex with VirFG<sup>N</sup>DD. Likewise, fused VirFG<sup>N</sup>DD incorporates two short  $\alpha$ -helices (residues A3–K11 and R17–L24 (numbered according to our revised sequence of VirFG<sup>21</sup>, GenBank: KT308172)), and its NMR spectrum superimposes on that of the domain in non-covalent complex with VirA<sup>C</sup>DD (**Supplementary Fig. 6l**). Together, the four  $\alpha$ -helices form a four  $\alpha$ -helical bundle in which the two sets of  $\alpha$ -helices are set at an angle of approximately 127° with respect to each other. Thus, unlike the Class 1 and 2 docking domains in which the upstream and downstream subunits each contribute two copies of the C- and N-terminal docking domains to a single docking complex<sup>15–17</sup>, the interaction between VirA and VirFG likely involves formation of two distinct docking complexes, each of which comprises one C- and one N-terminal docking domain. The docking interaction in *trans*-AT PKSs is therefore fundamentally different from that in *cis*-AT PKSs. It is also notable that a search of the PDB for homologues using both VAST<sup>38</sup> and DALI<sup>39</sup> did not reveal a comparable topology for any known interprotein complexes, consistent with previous findings that IDPs characteristically form unusual interactions<sup>40</sup>. Further support for the validity of the structure was provided by analysis by SAXS, as the curve predicted using CRY SOL<sup>41</sup> from the NMR structure of the VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD fusion protein superimposed nearly perfectly with the SAXS data obtained on this construct, as well as those of the non-covalent complex between the two domains (VirA<sup>C</sup>DD/VirFG<sup>N</sup>DD) ( $\chi^2 = 2.90$  and 2.68, respectively; **Fig. 3c,d, Supplementary Fig. 8d,e** and **Supplementary Table 4**). Furthermore, the SAXS curve predicted by CRY SOL for VirFG<sup>N</sup>DD alone if it adopted the same structure as in the complex, clearly does not correspond to the experimental data obtained on this construct ( $\chi^2 = 31.55$ ), proving that VirFG<sup>N</sup>DD undergoes a structural transition in the presence of its partner VirA<sup>C</sup>DD (**Supplementary Fig. 8b**).

Within the complex, the docking interface comprises  $\sim 730 \text{ \AA}^2$ , an extensive surface area given the limited number of participating residues (17 aa within a 4  $\text{\AA}$  radius). This too is consistent with VirFG<sup>N</sup>DD being an IDP, as such regions are typically able to achieve a greater interface area/residue than comparable ordered proteins<sup>42</sup>. If two such complexes form simultaneously *in vivo* at the VirA/VirFG junction, the total interaction surface area would be  $\sim 1460 \text{ \AA}^2$ , which is intermediate between Class 1 ( $\sim 1100 \text{ \AA}^2$ ) and Class 2 ( $\sim 1800 \text{ \AA}^2$ ) docking domains and consistent with the affinity

measured by ITC. The docking interface is hydrophobic, and comprises residues L6939, L6943, V6946, I6953, A6956, V6957 and L6960 of VirA<sup>C</sup>DD and residues A3, L7, F10, A20 and L24 of VirFG<sup>N</sup>DD (**Fig. 3e,f**). The contact surface between the two VirA<sup>C</sup>DD  $\alpha$ -helices includes L6942, A6945, L6951, A6956 and L6959, while that between the two VirFG<sup>N</sup>DD  $\alpha$ -helices comprises I6, L15, A19 and L23. Inspection of the sequence alignments of the two docking domain families (**Supplementary Fig. 1b,c** and **Supplementary Table 6**) shows that hydrophobic residues are relatively well-conserved at the corresponding positions. Three electrostatic interactions are also present at the interface (residues (VirA<sup>C</sup>DD/VirFG<sup>N</sup>DD): D6940/R17, N6949/K11 and E6954/K4).



**Figure 3 | Analysis of docking between VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (a) Sequence of the VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD fusion construct, with VirA<sup>C</sup>DD indicated in red, and VirFG<sup>N</sup>DD in blue. The N-terminal vector-derived residues and the added linker are shown in black. The numbering above the sequence corresponds to that in GenBank, while that below is based on the construct itself. (b) NMR structure of the covalent docking domain complex VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD color-coded as in a. (c and d) The strong correspondence between the SAXS curve for VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD predicted from its NMR structure (CRYSOLO<sup>41</sup>), and those obtained for (c) the non-covalent VirA<sup>C</sup>DD/VirFG<sup>N</sup>DD complex and (d) VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD, supports not only the validity of the NMR structure, but also the fact that covalently fusing the two DDs does not alter their interaction. (e) Hydrophobic residues at the docking interface within VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD. (f) Electrostatic interactions at the docking interface within VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD which were targeted for site-directed mutagenesis.

## Analysis of protein-protein interaction determinants by site-directed mutagenesis and ITC

As our docking model derives from the structure of the covalent complex of the two partners, we aimed to test its applicability to the discrete domains. For this, we mutated both polar and hydrophobic residues lying at the interface of the complex in order to evaluate their effects on the interaction affinity by ITC. Each mutant was initially analyzed by circular dichroism (CD) to monitor for structural changes induced by the mutations, as these would also have an effect on measured  $K_d$  values. In fact, the CD experiments revealed that VirFG<sup>NDD</sup> underwent architectural alterations when mutated, even gaining increasing secondary structure (**Supplementary Fig. 7e**). This observation is again consistent with VirFG<sup>NDD</sup> being an IDP, and that its disordered character needs to be actively maintained against mutations which can increase structural content<sup>43</sup>. On the other hand, this behavior complicated interpretation of ITC data obtained using mutant VirFG<sup>NDD</sup>, as we could not clearly distinguish between mutations which disrupted an important interaction with VirA<sup>CDD</sup> and those which simply impacted the ability of VirFG<sup>NDD</sup> to undergo induced structuration. We therefore focused our attention on the VirA<sup>CDD</sup> variants. Specifically, we analyzed the effect of exchanging D6940, N6949 and E6954 for alanine, as well as mutations which reversed the charge on residues D6940 and E6954 (changed to R and K respectively), and introduced charge at the N6949 position. We were unable to purify N6949A as well as an additional L6942A mutant due to poor solubility, but all other mutant proteins were obtained in acceptable yield, and far-UV CD analysis confirmed that their structures were not perturbed significantly relative to wild type VirA<sup>CDD</sup> (**Supplementary Fig. 7f**).

Due to solubility limitations, we could not reach binding saturation with many of the mutants, so the binding affinities determined by ITC are only rough estimates (the stated errors represent those from fitting of a single experiment) (**Supplementary Fig. 5f–k**). Nevertheless, the effect of the mutation in each case was clear. The D6940A and E6954A charge-removal mutations decreased binding affinity relative to that of wild type ( $K_d = 5.8 \mu\text{M}$ ) by factors of ca. 15 ( $85 \pm 23 \mu\text{M}$ ) and 4-fold ( $20 \pm 2 \mu\text{M}$ ), respectively (**Table 1**). The effect of changing the charge state of the residues was equally significant, with the affinity of VirFG<sup>NDD</sup> binding to the VirA<sup>CDD</sup> N6949K and E6954K mutants reduced to  $37 \pm 4 \mu\text{M}$  and  $49 \pm 16 \mu\text{M}$ , respectively. Even more dramatically, interaction between the D6940R mutant and VirFG<sup>NDD</sup> was completely abolished. Based on these results, we decided to test the effect of a combined mutation, introducing D6940A and E6954A simultaneously into VirA<sup>CDD</sup>. Analysis of the double mutant by CD showed that the structure was essentially unchanged, and the measured affinity to wild type VirFG<sup>NDD</sup> was  $>160 \mu\text{M}$ , demonstrating that the effect of the mutations was cumulative.

Taken together, our data clearly show that electrostatic interactions contribute to the assembly of the VirA<sup>CDD</sup>/VirFG<sup>NDD</sup> complex. Inspection of the residue positions corresponding to D6940, N6949 and E6954 in the other docking domains of the family shows that non-polar residues are also present in a number of systems, and that even when the residues are charged, they are not always matched by an opposite charge on the partner docking domain (**Supplementary Fig. 2b,c** and **Supplementary Table 6**). However in no case do the docking domain pairs from the same system have identical sets

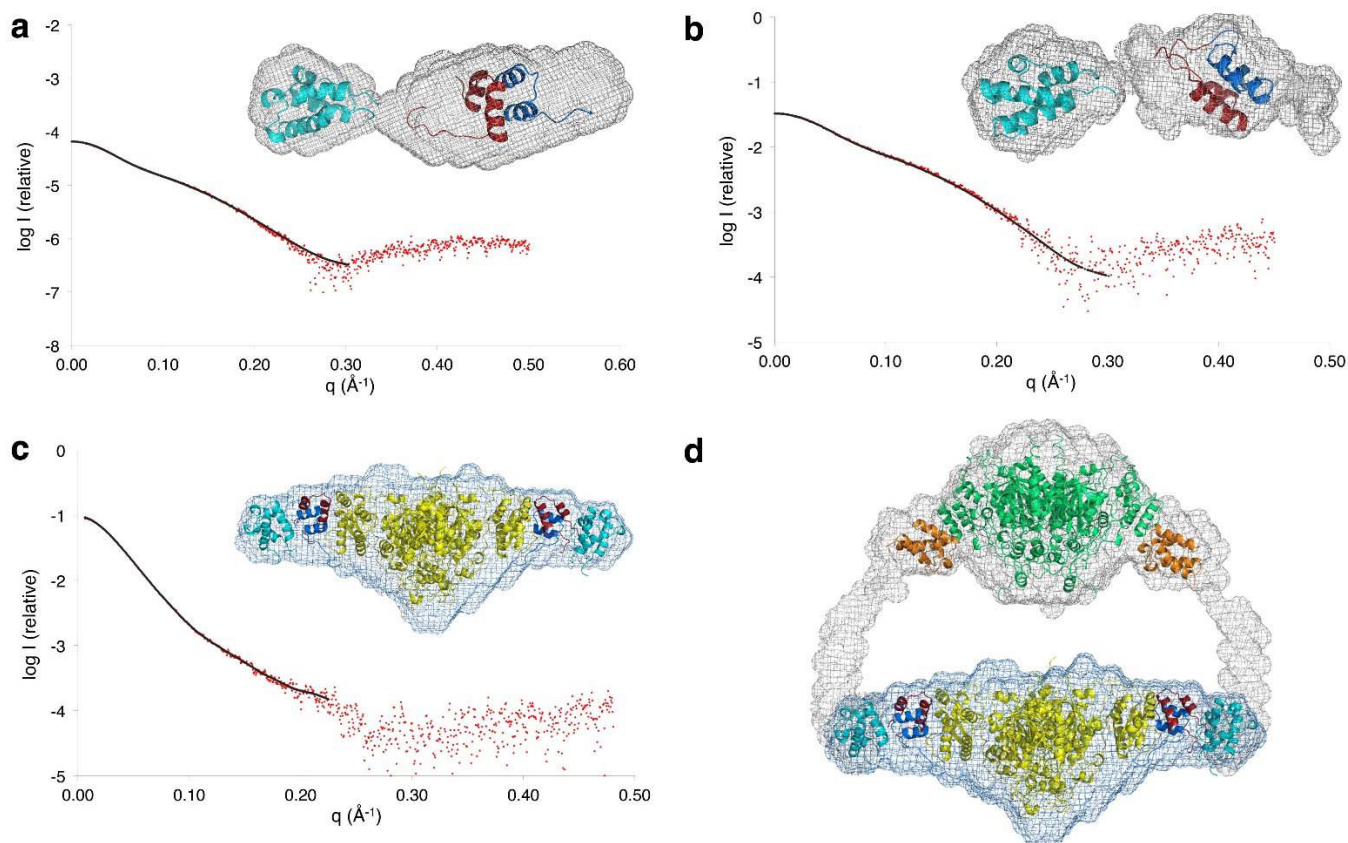
of residues at these positions, consistent with their potential contribution to interaction specificity through effects on  $k_{on}$ <sup>44,45</sup>. Similarly, the precise nature of the hydrophobic interface residues varies between multiple docking domain pairs from the same system (e.g. chivosazol (Chi), oxazolomycin (Ozm), thailandamide (Tai) and enacyloxin (5922)), and thus they may also contribute to docking specificity through effects on  $k_{on}$ ,  $k_{off}$  or both (**Supplementary Table 6**)<sup>44,45</sup>.

### Positioning of the flanking ACP and KS domains as determined by SAXS

The challenge of achieving specific interactions is substantial when multiple, candidate interaction partners have similar sequences and three-dimensional structures, as is the case within the C- and N-terminal docking domain families from the Chi, Ozm, enacyloxin and Tai systems (**Supplementary Table 2**)<sup>46</sup>. This raises the question of whether contacts between the flanking domains and the docking complex, or direct contact between the domains themselves, also contribute to specificity as observed for *cis*-AT PKSs<sup>24</sup>. To address this issue, we aimed to identify the locations of the flanking VirA ACP and VirFG KS domains relative to the docking domains, by small-angle X-ray scattering (SAXS). For this, we analyzed the following constructs: VirA ACP<sub>5b</sub>-<sup>C</sup>DD<sup>21</sup>, VirFG <sup>N</sup>DD-KS<sub>6</sub>, VirA ACP<sub>5b</sub>-<sup>C</sup>DD/VirFG <sup>N</sup>DD complex, VirA ACP<sub>5b</sub>-<sup>C</sup>DD-VirFG <sup>N</sup>DD (a covalent fusion between ACP<sub>5b</sub>-<sup>C</sup>DD and VirFG <sup>N</sup>DD), VirA <sup>C</sup>DD/VirFG <sup>N</sup>DD-KS<sub>6</sub> complex, and VirA ACP<sub>5b</sub>-<sup>C</sup>DD/VirFG <sup>N</sup>DD-KS<sub>6</sub> complex. For the non-covalent complexes, we used equal concentrations of both partners well above the measured  $K_d$  values in order to favor complex formation and disfavor self-association of VirA <sup>C</sup>DD. Unfortunately, VirFG <sup>N</sup>DD-KS<sub>6</sub> proved to readily self-aggregate, so for these experiments we prepared a new fusion protein spanning from ACP<sub>5b</sub> to KS<sub>6</sub> (VirA ACP<sub>5b</sub>-<sup>C</sup>DD-VirFG <sup>N</sup>DD-KS<sub>6</sub>) (**Supplementary Fig. 3**).

Previous analysis of VirA ACP<sub>5b</sub>-<sup>C</sup>DD yielded an average molecular envelope consistent with the homodimerization observed by NMR and ITC<sup>21</sup>. In the presence of equimolar concentrations of its partner VirFG <sup>N</sup>DD, however, a new average molecular envelope was obtained (**Fig. 4a**, **Supplementary Fig. 8f** and **Supplementary Table 4**), in which density was observed only for a single ACP domain and the docking domain complex. Essentially identical data were acquired on VirA ACP<sub>5b</sub>-<sup>C</sup>DD-VirFG <sup>N</sup>DD (**Fig. 4b** and **Supplementary Fig. 8g**), strengthening the conclusion that joining the docking domains covalently does not alter how they interact. The data also confirm that the docking complex is formed from only a single copy each of the C- and N-terminal docking domains, and that ACP<sub>5b</sub> does not contact it directly. Analysis of VirA ACP<sub>5b</sub>-<sup>C</sup>DD-VirFG <sup>N</sup>DD-KS<sub>6</sub> (**Supplementary Fig. 8h**) gave an average molecule envelope ( $\chi^2 = 1.77$ ) into which the homodimeric KS<sub>6</sub> domain could be confidently placed using SUPCOMB<sup>47</sup> (**Fig. 4c**). It was not possible, however, to computationally locate the two ACP<sub>5b</sub> domains and the two docking domain complexes because the scattering data are dominated by the signal from the KS<sub>6</sub>. However, from the obtained molecular envelope it is clear that the 42-residue sequence separating VirFG <sup>N</sup>DD from KS<sub>6</sub> must be largely structured, as there is no electronic density which could correspond to an unstructured region of this size. Furthermore, making the reasonable assumption that the ACP<sub>5b</sub> domains are located on the two extremities of the structure (**Fig. 4c**) allowed us to superimpose the molecular envelope on that obtained previously for the portion of module 5 encompassing the KS<sub>5</sub>, ACP<sub>5a</sub> and ACP<sub>5b</sub> domains (KS<sub>5</sub>-ACP<sub>5a</sub>-ACP<sub>5b</sub>)<sup>21</sup> (**Fig. 4d**). This analysis shows that the two sets of ACP<sub>5b</sub> domains superimpose perfectly and therefore that the global form observed for VirA ACP<sub>5b</sub>-

${}^{\text{C}}\text{DD-VirFG}^{\text{N}}\text{DD-KS}_6$  is compatible with the previously-characterized structure of module 5 (ref. 21). In the absence of docking with the downstream subunit, module 5 adopts a wider open shape (distance  $\text{ACP}_{5b}$  to  $\text{ACP}_{5b}$  in module 5 alone =  $279 \text{ \AA}^{21}$  vs.  $200 \text{ \AA}$  in  $\text{KS}_5\text{-ACP}_{5a}\text{-ACP}_{5b}$ ), and so, as previously proposed<sup>21</sup>, interaction with VirFG induces closure of the module.



**Figure 4 | Analysis of the interface between subunits VirA and VirFG by SAXS.** (a) Analysis by SAXS of the VirA  $\text{ACP}_{5b}\text{-}^{\text{C}}\text{DD}/\text{VirFG}^{\text{N}}\text{DD}$  complex. Fit between the *ab initio* model computed with DAMMIN<sup>30</sup> (solid black line) and the experimental SAXS data (red dots). This analysis yielded a molecular envelope different from that of VirA  $\text{ACP}_{5b}\text{-}^{\text{C}}\text{DD}$  alone<sup>21</sup>, in which density was observed for one ACP domain (light blue) and the docking domain complex (VirA  ${}^{\text{C}}\text{DD}$  is shown in red and VirFG  ${}^{\text{N}}\text{DD}$  in dark blue). This result confirmed that docking occurs between a single copy each of the C- and N-terminal docking domains, and that  $\text{ACP}_{5b}$  does not directly contact the complex of DDs. (b) Analysis of VirA  $\text{ACP}_{5b}\text{-}^{\text{C}}\text{DD-VirFG}^{\text{N}}\text{DD}$  by SAXS gave similar data to that shown in a. (Mention the difference can be explained by the covalent link) Fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots). (c) Analysis by SAXS of a covalent complex between VirA  $\text{ACP}_{5b}\text{-}^{\text{C}}\text{DD}$  and VirFG  ${}^{\text{N}}\text{DD-KS}_6$  (VirA  $\text{ACP}_{5b}\text{-}^{\text{C}}\text{DD-VirFG}^{\text{N}}\text{DD-KS}_6$ ). Fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots). A homology model of  $\text{KS}_6$  (yellow) was placed into the average molecular envelope using SUPCOMB<sup>47</sup>, while localization of the NMR structure of  $\text{ACP}_{5b}$ <sup>21</sup> was based on the assumption that the domains would be situated on the extremities of the structure. The remaining electronic density could then be assigned to the docking domain complex. (d) Model for the complete interface between subunits VirA and VirFG generated by superimposing the SAXS-derived model of the  $\text{KS}_5\text{-ACP}_{5a}\text{-ACP}_{5b}$  region of VirA module 5 (ref. 21) ( $\text{KS}_5$  in green,  $\text{ACP}_{5a}$  in orange) with the average molecular envelope obtained for VirA  $\text{ACP}_{5b}\text{-}^{\text{C}}\text{DD-VirFG}^{\text{N}}\text{DD-KS}_6$ . This analysis suggests that the interaction between the subunits creates a single, closed reaction chamber within module 5. However, as the estimated distance between the  $\text{ACP}_{5b}$  and  $\text{KS}_6$  active sites would not allow for direct chain transfer, conformational adjustment is likely to occur in the presence of polyketide intermediate on  $\text{ACP}_{5b}$ <sup>7</sup>.

## DISCUSSION

The majority of modular PKSs incorporate multiple polypeptide subunits, suggesting an evolutionary advantage in distributing the modules among several smaller proteins. Communication between subunits in all *cis*-AT PKSs characterized to date is orchestrated, at least in part, by matched terminal docking domains (DDs)<sup>14-17</sup>. We have shown here that for an important subset of interpolypeptide junctions, analogous DDs operate in *trans*-AT PKSs. The recognition of four distinct structural classes of docking domains now permits a systematic comparison of docking strategies which



give rise to both high specificity but relatively low affinity interactions (ca. 2–50  $\mu\text{M}$ ). All of the docking complexes incorporate contacts between  $\alpha$ -helices. Although architecturally distinct, the docking interfaces are largely hydrophobic and similar in size (1100–1800  $\text{\AA}^2$ ). Selectivity among potential partners seems to arise in all cases from electrostatic interactions between a few, strategically placed charged residues, although many PKS assembly lines also incorporate several, structurally orthogonal classes of DDs (i.e. classes 1a and 1b<sup>13</sup>).

We have provided evidence that the weak docking affinity arises in the case of multiple *trans*-AT PKSs through use of an intrinsically disordered protein as a docking partner. IDPs are rare in prokaryotes<sup>48</sup> (though relatively more common in high GC, large genome-size bacteria such as the Actinobacteria<sup>49</sup>) and VirFG<sup>NDD</sup> represents, to our knowledge, the first IDP to be shown to participate in bacterial secondary metabolism. VirA<sup>CDD</sup> may also be an IDP, but a disorder to order transition could not be observed in the context of a discrete protein because of self-association. IDPs are expected to show much higher rates of sequence variation than structured regions of the same proteins<sup>43</sup>. This is indeed the case for the Vir<sup>4</sup> and Sna<sup>26</sup> systems which both govern virginiamycin biosynthesis, as the corresponding subunits show much higher amino acid sequence identity than the docking domain regions (e.g. 66% identity for VirA/SnaE2, but only 53% for the CDDs; 66% overall identity between VirFG/SnaE3, vs. 39% for the NDDs<sup>50</sup>). This observation suggests that it is the amino acid character and not the specific sequences of these regions that is critical. However, given the unpredictable effects of mutations on induced folding and structure of IDPs<sup>43</sup>, it seems likely that using this class of DD to engineer new interfaces will require transplantation of matched sets of docking domains. Our study identifies clear boundaries to be used in such experiments.

Our work has also permitted for the first time visualization of a full interface formed between two consecutive subunits within a *trans*-AT PKS. Our previous analysis by SAXS showed VirA module 5 to have an open architecture, but that flexibility on the level of the ‘arms’ might allow it to clamp down to engage with the KS<sub>6</sub> domain of the next module. Indeed, the distance between the ACP<sub>5b</sub> domains seems perfectly adapted to accommodate the homodimeric KS<sub>6</sub> and two flanking docking domain complexes. In this way, docking between the two subunits creates a closed space containing a single reaction center in which reactive intermediates could be protected. On the other hand, the estimated distance between ACP<sub>5b</sub> and KS<sub>6</sub> active sites in this *apo* structure (ca. 80  $\text{\AA}$ ) is not compatible with direct chain transfer between the two domains as required by the biosynthetic mechanism. This observation is consistent with recent studies by cryo-electron microscopy (cryo-EM) on the interface between subunits from the pikromycin (PIK) *cis*-AT PKS, in which a stable interaction between the partner ACP and KS domains was only observed in the presence of chain extension intermediate tethered to the ACP<sup>6</sup>. We propose that the intrinsic flexibility of the linker regions separating the various functional and docking domains will allow adoption of a configuration in which ACP<sub>5b</sub> is correctly positioned to interact with KS<sub>6</sub>. However, visualizing this productive chain transfer complex by SAXS or cryo-EM will likely require studies of appropriately modified VirA module 5, work which is in progress. In the meantime, the absence of a specific interface between the ACP and KS domains is encouraging from an engineering perspective, because it suggests that transplantation of docking domains may be sufficient to create novel productive intersubunit interfaces in *trans*-AT PKSs, as long as the receiving KS has appropriate substrate specificity<sup>27</sup>.

## METHODS

Methods and any associated references are available in the online version of the paper.

**Accession codes.** Chemical shifts for VirA<sup>C</sup>DD in complex with VirFG<sup>N</sup>DD and for VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD have been deposited in the Biological Magnetic Resonance Bank with accession codes 25721 and 25706, respectively. Protein Data Bank coordinates for VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD have been deposited under accession code 2N5D.

## References

1. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed Engl.* **48**, 4688–4716 (2009).
2. Katz, L. The DEBS paradigm for type I modular polyketide synthases and beyond. *Methods Enzymol.* **459**, 113–142 (2009).
3. Piel, J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2010).
4. Pulsawat, N., Kitani, S. & Nihira, T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **393**, 31–42 (2007).
5. Weissman, K.J. & Müller, R. Protein–protein interactions in multienzyme megasynthetases. *ChemBioChem* **9**, 826–848 (2008).
6. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
7. Whicher, J.R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).
8. Gokhale, R.S., Tsuji, S.Y., Cane, D.E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
9. Tsuji, S.Y., Cane, D.E. & Khosla, C. Selective protein–protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **40**, 2326–2331 (2001).
10. Menzella, H.G. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).
11. Menzella, H.G., Carney, J.R. & Santi, D.V. Rational design and assembly of synthetic trimodular polyketide synthases. *Chem. Biol.* **14**, 143–151 (2007).
12. Reeves, C.D. *et al.* Production of hybrid 16-membered macrolides by expressing combinations of polyketide synthase genes in engineered *Streptomyces fradiae* hosts. *Chem. Biol.* **11**, 1465–1472 (2004).
13. Thattai, M., Burak, Y. & Shraiman, B.I. The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **3**, 1827–1835 (2007).
14. Richter, C.D., Nietlispach, D., Broadhurst, R.W. & Weissman, K.J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2008).
15. Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F. & Weissman, K.J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
16. Buchholz, T.J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).

17. Whicher, J.R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
18. Tang, Y., Kim, C.-Y., Mathews, I.I., Cane, D.E. & Khosla, C. The 2.7 Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–11129 (2006).
19. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D.E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931–943 (2007).
20. Gokhale, R.S., Hunziker, D., Cane, D.E. & Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chem. Biol.* **6**, 117–125 (1999).
21. Davison, J. *et al.* Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem. Sci.* **5**, 3081–3095 (2014).
22. Gu, L. *et al.* Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect. *Angew. Chem. Int. Ed Engl.* **50**, 2795–2798 (2011).
23. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Wu, N., Cane, D.E. & Khosla, C. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41**, 5056–5066 (2002).
25. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
26. Mast, Y. *et al.* Characterization of the ‘pristinamycin supercluster’ of *Streptomyces pristinaespiralis*. *Microb. Biotechnol.* **4**, 192–206 (2011).
27. Nguyen, T. *et al.* Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
28. Sibille, N. & Bernadó, P. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.* **40**, 955–962 (2012).
29. Bourhis, J.-M. *et al.* The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci.* **14**, 1975–1992 (2005).
30. Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–86 (1999).
31. Receveur-Brechot, V. & Durand, D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13**, 55–75 (2012).
32. Böhm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5**, 191–195 (1992).
33. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).
34. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
35. Case, D.A. *et al.* AMBER 14, University of California, San Francisco.
36. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).

37. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
38. Bailey, T.L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998).
39. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549 (2010).
40. Uversky, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* **22**, 693–724 (2013).
41. Svergun, D., Barberato, C. & Koch, M. CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773
42. Liu, Z. & Huang, Y. Advantages of proteins being disordered. *Protein Sci.* **23**, 539–550 (2014).
43. Oldfield, C.J. & Dunker, A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
44. Selzer, T. & Schreiber, G. New insights into the mechanism of protein-protein association. *Proteins* **45**, 190–198 (2001).
45. Schreiber, G., Haran, G. & Zhou, H.–X. Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* **109**, 839–860 (2009).
46. Schreiber, G. & Keating, A.E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).
47. M Kozin & Svergun, D. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41
48. Williamson, M.P. & Potts, J.R. Intrinsically disordered proteins: administration not executive. *Biochem. Soc. Trans.* **40**, 945–949 (2012).
49. Pavlović–Lažetić, G.M. *et al.* Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics* **12**, 66 (2011).
50. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
51. Keating, T.A., Marshall, C.G., Walsh, C.T. & Keating, A.E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
52. Zheng, J., Gay, D.C., Demeler, B., White, M.A. & Keatinge–Clay, A.T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615–621 (2012).
53. *Molecular Cloning: A Laboratory Manual*, Eds.: J. Sambrook, E. F. Fritsch, T. Maniatis (Cold Spring Harbor Laboratory Press, New York, 1989).
54. Keller, S. *et al.* High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* **84**, 5066–5073 (2012).
55. Cavanagh, J., Fairbrother, W., Palmer, A. & Skelton, N. *Protein NMR Spectroscopy: Principles and Practice* (Academic Press Inc., New York, 1996).
56. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227 (2002).
57. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
58. David, G. & Pérez, J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Crystallogr.* **42**, 892–900

59. Pieper, R., Ebert-Khosla, S., Cane, D. & Khosla, C. Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates. *Biochemistry* **35**, 2054–2060 (1996).
60. Bycroft, M., Weissman, K.J., Staunton, J. & Leadlay, P.F. Efficient purification and kinetic characterization of a bimodular derivative of the erythromycin polyketide synthase. *Eur. J. Biochem.* **267**, 520–526 (2000).
61. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
62. Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
63. Volkov, V. & Svergun, D. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
64. Söding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–248 (2005).
65. Gay, D.C. *et al.* A close look at a ketosynthase from a *trans*-acyltransferase modular polyketide synthase. *Structure* **22**, 444–451 (2014).
66. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* Chapter 2, Unit 2.9 (2007).

## Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR JCJC 2011 PKS-PPIs to K.J.W.), the Centre National de la Recherche Scientifique (CNRS), the Université de Lorraine (UL), and the Région Lorraine Region (Bonus Qualité Recherche (BQR) grants to K.J.W. and B.C). Access to the 600 MHz NMR spectrometer of the ‘SCBIM’ (Service Commun de Biophysicochimie des Interactions Moléculaires) of the FR3209 CNRS/Université de Lorraine was greatly appreciated. We thank Dr Javier Perez and Dr Aurélien Thureau for help during data collection on the beamline SWING at the synchrotron SOLEIL (Gif-sur-Yvette). We would also like to acknowledge L. Ornella Bimai for help in preparing the docking domain constructs and Alexandre Kriznik for assistance in analyzing the ITC and CD data.

## Author contributions

B.C., A.G. and K.J.W. designed the study. K.J.W. carried out the comparative docking domain sequence analysis. T.A. and J.D. cloned, expressed and purified the recombinant protein constructs with the help of F.R., S.C. and C.J. F.R. carried out the analyses by CD and ITC under supervision by T.A., J.D. and B.C. T.A. and J.D. carried out the NMR and SAXS experiments with B.C. and A.G., respectively. T.A., J.D., F.R., A.G., B.C. and K.J.W. interpreted the data. K.J.W. wrote the paper, with input from T.A., J.D., F.R., C.J., A.G. and B.C.

## Competing financial interests

The authors declare no competing financial interests.

## Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to A.G., B.C. or K.J.W.

**Table 1 | Effect of VirA<sup>C</sup>DD point mutations on the affinity of docking.**

VirA <sup>C</sup> DD construct	Affinity for wild type VirFG <sup>N</sup> DD (μM)
Wild type	5.8 ± 0.2 <sup>a</sup>
D6940A	85 ± 23
D6940R	No binding
L6942A	n.d. <sup>b</sup>
N6949A	n.d. <sup>b</sup>
N6949K	37 ± 4
E6954A	20 ± 2
E6954K	49 ± 16
D6940A + E6954A	>160

<sup>a</sup>Average of three independent measurements; all others represent the error on a single measurement

<sup>b</sup>Not determined as the constructs were poorly soluble

## ONLINE METHODS

**Sequence analysis.** The sequences of the analyzed *trans*-AT PKSs were accessed via PubMed, or in the case of the class 2 C- and N-terminal docking domains, from ref. 17. Sequence alignments were generated using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), and displayed using BoxShade ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)). The C-terminal boundaries of the ACP domains were determined by multiple sequence alignment of the subunit ends, and comparison with the solved structures of ACPs 5a and 5b from the virginiamycin PKS<sup>21</sup>. Similarly, the start sites of the domains downstream of the N-terminal docking domains were assigned by alignment of the beginning of the subunits and comparison with the sequences of the solved structures of KS<sup>18,19</sup>, C<sup>51</sup> and ER<sup>52</sup> domains. Secondary structure prediction was carried out with PsiPred<sup>25</sup>.

**Materials, DNA manipulation and sequencing.** Biochemicals and media were purchased from ThermoFisher Scientific (Tris, EDTA), Merck (NaPi), Carlo ERBA (NaCl), BD (peptone, yeast extract), VWR (glycerol) and Sigma-Aldrich (imidazole, IPTG). The enzymes for genetic manipulation were purchased from ThermoFisher Scientific. DNA isolation and manipulation in *E. coli* was carried out according to standard methods<sup>53</sup>. PCR amplifications were carried out on a Mastercycler Pro (Eppendorf) using Phusion high-fidelity DNA polymerase (Thermo Scientific). Primers were obtained from Sigma-Aldrich (France) and DNA sequencing was carried out by GATC Biotech (Mulhouse, France).

**Cloning, expression and purification.** Five constructs (VirA<sup>C</sup>DD, VirFG<sup>N</sup>DD, VirFG<sup>N</sup>DD-linker, VirA ACP<sub>5b</sub><sup>C</sup>DD and VirFG<sup>N</sup>DD-KS<sub>6</sub>) were amplified from *Streptomyces virginiae* genomic DNA using forward and reverse primers incorporating *Bam*HI and *Hind*III restriction sites, respectively (see **Supplementary Methods, Supplementary Table 7**), and were ligated into the equivalent sites of vector pBG-102; in the case of VirFG<sup>N</sup>DD-KS<sub>6</sub> an internal *Bam*HI site was removed by site-directed mutagenesis using Phusion high-fidelity DNA polymerase, without introducing changes in the amino acid sequence. Vector pBG-102 codes for a SUMO-His<sub>6</sub> tag (Center for Structural Biology, Vanderbilt University). Following cleavage of the SUMO-His<sub>6</sub> tag, the proteins incorporated a non-native N-terminal Gly–Pro–Gly–Ser sequence. From these vectors and *S. virginiae* genomic DNA, covalent fusions (VirA<sup>C</sup>DD–VirFG<sup>N</sup>DD, VirA ACP<sub>5b</sub><sup>C</sup>DD–VirFG<sup>N</sup>DD and VirA ACP<sub>5b</sub><sup>C</sup>DD–VirFG<sup>N</sup>DD-KS<sub>6</sub>) were constructed using overlap PCR. Fragments encoding VirA<sup>C</sup>DD, VirFG<sup>N</sup>DD, VirA ACP<sub>5b</sub><sup>C</sup>DD and VirFG<sup>N</sup>DD-KS<sub>6</sub> were amplified and then used as templates in a second PCR to amplify the fusion construct, and the resulting fragments ligated into the vector pBG-102. Site-directed mutants of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD were obtained by site-directed mutagenesis using Phusion high-fidelity DNA polymerase from vectors pBG-102 containing VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD wild type sequences.

The vectors were used to transform *E. coli* BL21 (DE3), and constructs were expressed as SUMO-His<sub>6</sub>-tagged fusions by growth in LB medium at 37 °C to an A<sub>600</sub> of 0.8, followed by induction with IPTG (0.1 mM) and incubation at 20 °C for a further 12–18 h. The *E. coli* cells were collected by centrifugation and resuspended in buffer 1 (50 mM sodium phosphate (pH 7.5), 250 mM NaCl) in the case of VirA<sup>C</sup>DD, VirA ACP<sub>5b</sub><sup>C</sup>DD, VirA ACP<sub>5b</sub><sup>C</sup>DD–VirFG<sup>N</sup>DD (covalent fusion), and all VirA

<sup>C</sup>DD mutants, buffer 2 (50 mM Tris-HCl (pH 7.5), 150 mM NaCl) in the case of VirFG <sup>N</sup>DD, VirFG <sup>N</sup>DD-linker, VirA <sup>C</sup>DD–VirFG <sup>N</sup>DD (covalent fusion) and all VirFG <sup>N</sup>DD mutants, and in buffer 3 (50 mM sodium phosphate (pH 7.5), 400 mM NaCl, 10% glycerol) for VirA ACP<sub>5b</sub>-<sup>C</sup>DD–VirFG <sup>N</sup>DD-KS<sub>6</sub> (covalent fusion). Cells were lysed by sonication, and cell debris were removed by centrifugation and filtration (0.45 μm). The cell lysates were loaded onto a HisTrap 5 ml column (GE), equilibrated in buffer 1, 2 or 3 as appropriate. The column was washed extensively with buffer 1, 2 or 3 containing 60 mM imidazole, and His-tagged proteins were eluted using 300 mM imidazole. The VirA ACP<sub>5b</sub>-<sup>C</sup>DD–VirFG <sup>N</sup>DD-KS<sub>6</sub> covalent fusion was further loaded onto a HiTrap Q HP column (GE) equilibrated in buffer (20 mM sodium phosphate (pH 7.5), 2 mM EDTA, 5% glycerol) which was then washed extensively with this buffer. The fusion was eluted using a NaCl gradient, up to a final concentration of 1 M. All SUMO-His<sub>6</sub> tagged constructs were then incubated with His-tagged human rhinovirus 3C protease (1 mM) for 12–16 h at 4 °C to cleave the affinity-solubility tags. The constructs were then separated from the remaining His-tagged proteins by loading onto a HisTrap 5 ml column (GE) followed by elution in buffer 1, 2 or 3 containing 20 mM imidazole.

Following this step, size exclusion chromatography on a Superdex 75 26/60 column (GE) in 100 mM sodium phosphate (pH 7.5) or on a Superdex 200 26/60 column (GE) in the case of VirA ACP<sub>5b</sub>-<sup>C</sup>DD–VirFG <sup>N</sup>DD-KS<sub>6</sub> (covalent fusion) in 20 mM sodium phosphate (pH 7.5), 250 mM NaCl, 2 mM EDTA, 5% glycerol, resulted in a homogeneous preparation of each protein. Production of <sup>15</sup>N or <sup>13</sup>C,<sup>15</sup>N-enriched VirA <sup>C</sup>DD, VirFG <sup>N</sup>DD and VirA <sup>C</sup>DD–VirFG <sup>N</sup>DD (covalent fusion) for NMR structure elucidation was carried out by growth in M9 minimal medium. The minimal medium was supplemented with <sup>15</sup>NH<sub>4</sub>Cl (0.5 g/l) and glucose <sup>13</sup>C<sub>6</sub>-glucose (2.0 g/l) as the sole sources of nitrogen and carbon.

**Analytical gel filtration.** VirA <sup>C</sup>DD and VirFG <sup>N</sup>DD (initial concentrations of 670 μM) were loaded separately onto a Superdex 75 10/300 GL column (GE) equilibrated with 100 mM sodium phosphate (pH 6.5) buffer and elution of the two proteins followed via their absorbance at 280 nm. The same volume of a mixture of the two proteins (both at 670 μM) was then re-injected.

**Isothermal titration calorimetry.** ITC was performed using a MicroCal ITC 200 (GE) at 20 °C. 300 μL of VirA <sup>C</sup>DD (wild type or mutant) at 60 μM were placed in the calorimeter cell, and VirFG <sup>N</sup>DD (wild type or mutant at 600 μM) or VirFG <sup>N</sup>DD-linker (at 1.4 mM) were added as follows: 0.4 μL over 0.8 s for the first injection, followed by 19 injections of 2 μL over 4 s at 120 s intervals. For the experiment with VirA ACP<sub>5b</sub>-<sup>C</sup>DD, the construct (900 μM) was added to VirFG <sup>N</sup>DD (60 μM), as described above. The heat of reaction per injection (micro-calories/second) was determined by integration of the peak areas using either the Origin (MicroCal) software or Nitpic<sup>54</sup>. These software provide the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N), and the dissociation constant (K<sub>d</sub>). The heats of dilution were determined by injecting VirA <sup>C</sup>DD and VirFG <sup>N</sup>DD alone into the buffer (initial concentrations of 600 μM), and were subtracted from the corresponding experiments before curve-fitting. This experiment additionally showed that VirA <sup>C</sup>DD was undergoing homodimerization as the concentration in the calorimeter cell increased, as a change in the heat signal was detected up



to a concentration of 30  $\mu\text{M}$ . No such change was observed over the whole concentration range for VirFG  $^{\text{N}}\text{DD}$  (up to 78  $\mu\text{M}$ ).

**NMR data acquisition.** To minimize the amount of protein needed, 300  $\mu\text{L}$  of protein solution were loaded into a 4 mm NMR tube. [ $^1\text{H}$ - $^{15}\text{N}$ ]-HSQC spectra of VirA  $^{\text{C}}\text{DD}$  were recorded at a concentration of 100  $\mu\text{M}$  in the absence of partner, and in the presence of 200–400  $\mu\text{M}$   $^{14}\text{N}$  VirFG  $^{\text{N}}\text{DD}$ . [ $^1\text{H}$ - $^{15}\text{N}$ ]-HSQC spectra of VirFG  $^{\text{N}}\text{DD}$  were recorded at a concentration of 100  $\mu\text{M}$  either alone, or in the presence of 400  $\mu\text{M}$   $^{14}\text{N}$  VirA  $^{\text{C}}\text{DD}$  or 200  $\mu\text{M}$   $^{14}\text{N}$  VirA ACP<sub>5b</sub>- $^{\text{C}}\text{DD}$ . Heteronuclear spectra were acquired on  $^{13}\text{C}$ , $^{15}\text{N}$ -VirA  $^{\text{C}}\text{DD}$  at a concentration of 1.2 mM in the presence of a 2-fold excess of  $^{12}\text{C}$ , $^{14}\text{N}$ -VirFG, and on VirA  $^{\text{C}}\text{DD}$ -VirFG  $^{\text{N}}\text{DD}$  at a concentration of 0.9 mM. All NMR data were recorded at 298 K on a Bruker DRX600 spectrometer equipped with a cryogenic probe (Plateforme de Biophysico-chimie Biologie Structurale – FR 3209 Université de Lorraine/CNRS).

Backbone and sequential resonance assignments were obtained by the combined use of 2D  $^{15}\text{N}$ - $^1\text{H}$  and  $^{13}\text{C}$ - $^1\text{H}$  HSQC spectra and 3D HNCACB, CBCA(CO)NH and HNCOC experiments (reviewed in ref. 55). Assignments of aliphatic side chain resonance were based on 3D (H)CC(CO)NH, H(CCCO)NH, and HCCH-TOCSY experiments. To collect NOE-based distance restraints for the VirA  $^{\text{C}}\text{DD}$ -VirFG  $^{\text{N}}\text{DD}$  structure calculations, 3D  $^{15}\text{N}$  NOESY-HSQC and  $^{13}\text{C}$  NOESY-HSQC were recorded on uniformly  $^{13}\text{C}$ , $^{15}\text{N}$ -enriched VirA  $^{\text{C}}\text{DD}$ -VirFG  $^{\text{N}}\text{DD}$  samples. The mixing time used in all NOESY experiments was set to 120 ms. NMR data were processed using Topspin 3.1 (Bruker) and were analyzed using Sparky (Goddard TD and Kneller DG, SPARKY3, University of California, San Francisco, 2003).

**NMR structure calculation.** Initial structures were generated using CYANA 3.0 software<sup>33</sup>. Starting from a set of manually assigned NOEs, the standard CYANA protocol of seven iterative cycles of calculations was performed with NOE assignment by the embedded CANDID routine combined with torsion angle dynamics structure calculation<sup>56</sup>. In each cycle, 100 structures starting from random torsion angle values were calculated with 10,000 steps of torsion angle dynamics-driven simulated annealing. 478 NOE-based distance and 54 backbone angle restraints were used for structure calculation of the VirA  $^{\text{C}}\text{DD}$  domain, while 1372 NOE-based distance and 98 backbone angle restraints were used for VirA  $^{\text{C}}\text{DD}$ -VirFG  $^{\text{N}}\text{DD}$ . The angle restraints were obtained from  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$ ,  $^{13}\text{C}'$ ,  $^{15}\text{N}$ ,  $^1\text{HN}$  and  $^1\text{H}\alpha$  chemical shifts using TALOS+<sup>57</sup> with an assigned minimum range of  $\pm 20^\circ$ . No hydrogen-bond restraints were used for structure calculation. The second stage consisted of the refinement of the 50 lowest CYANA target function conformers by restrained molecular dynamic (rMD) simulations in Amber 14<sup>35</sup> using the generalized Born solvent model. Possible atoms clashes within CYANA initial structures were regularized by a 1-ps energy minimization in the Amber force field, and then 20 ps of rMD was performed with the following protocol: the system was heated for 5 ps at 600 K with tight coupling for heating and equilibration (TAUTP = 0.4), 15 ps of cooling to 0 K with 13 ps of slow cooling (loose coupling, TAUTP = 4.0–1.0) followed by 1 ps of faster cooling (TAUTP = 1.0), and a final 1 ps of very fast cooling (TAUTP = 0.1–0.05). During rMD, the restraints were slowly ramped from 10–100% of their final values over the first 3 ps. Force constants for distance and angle restraints were set to 32 kcal

$\text{mol}^{-1} \text{Å}^{-2}$  and  $50 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ , respectively. The final VirA<sup>CDD</sup>–VirFG<sup>NDD</sup> representative ensembles correspond to the 20 conformers from each calculation with the lowest restraint and Amber energy terms.

**Analysis by circular dichroism (CD).** CD was carried out on a Chirascan CD from Applied Photophysics. Data were collected at 0.5 nm intervals in the wavelength range of 180–260 nm at 20 °C, using a temperature-controlled chamber. A 0.01 cm cuvette containing 30  $\mu\text{L}$  of protein sample at 100  $\mu\text{M}$  was used for all the measurements except that of 10  $\mu\text{M}$  VirA<sup>CDD</sup> (100  $\mu\text{L}$  of sample were used in a 0.1 cm cuvette). Measurements were made at least in triplicate, and sample spectra were corrected for buffer background by subtracting the average spectrum of buffer alone. In order to evaluate the extent of induced folding when the two docking domains were together, we acquired a combined spectrum of the two docking domains by placing VirA<sup>CDD</sup> into one cuvette and VirFG<sup>NDD</sup> into another cuvette (30  $\mu\text{L}$  each at 100  $\mu\text{M}$ ), and analyzing the two cuvettes simultaneously. The same result was obtained by averaging the spectra of VirA<sup>CDD</sup> and VirFG<sup>NDD</sup> acquired separately. We then combined the two docking domains, and placed the mixture into both cuvettes (30  $\mu\text{L}$  at 100  $\mu\text{M}$ ), again analyzing the two simultaneously. Spectrum deconvolution was carried out using the CDNN software<sup>32</sup>.

**Small angle X-ray scattering (SAXS) data collection.** The SAXS data were collected on the SWING beamline at the Synchrotron SOLEIL (France). Frames were recorded using an AVIEX170170 CCD detector at the energy of 12 keV. The distance between the sample and the detector was set to 1803 mm, leading to scattering vectors  $q$  ranging from 0.0005  $\text{Å}^{-1}$  to 0.5  $\text{Å}^{-1}$ . The scattering vector is defined as  $4\pi/\lambda \sin\theta$ , where  $2\theta$  is the scattering angle. The protein samples were injected using the online automatic sample changer into a pre-equilibrated HPLC-coupled size-exclusion chromatography column (Bio-SEC 100, Bio-SEC 150 or Bio-SEC 300, Agilent) depending on the molecular weight of the protein sample, at a temperature of 15 °C. The online purification system that delivers the eluted fractions into the measurement cell was developed on the SWING beamline<sup>58</sup>. After equilibrating the column in the protein buffer (50 mM sodium phosphate (pH 7.5), 150 mM NaCl, 5% glycerol (as radioprotectant)), 100 successive frames of 0.75 s were recorded of the buffer background. This buffer was chosen for the SAXS analysis as phosphate, salt and glycerol are known to maintain the catalytic activity of *cis*-AT PKSs<sup>59,60</sup>, presumably by stabilizing the structure the proteins. 50  $\mu\text{L}$  of the protein sample (at 10 or 15 mg/ml) were then injected and complete data sets acquired. The protein concentration downstream of the elution column was followed via the absorbance at 280 nm with an *in situ* spectrophotometer. Recorded frames were then radially corrected, averaged and normalized using the program Foxtrot (in-house application). Data obtained on the fractions from the elution peaks were checked for stable radii of gyration, processed, corrected for buffer signal, and averaged using the Foxtrot program. Data processing was performed using the program package PRIMUS<sup>61</sup>. The radius of gyration  $R_g$  was derived by the Guinier approximation  $I(q) = I(0) \exp(-q^2 R_g^2/3)$  for  $qR_g < 1.3$ . The distance distribution function  $P(r)$  and the maximum particle diameter  $D_{max}$  (Supplementary Table 4) were calculated by Fourier inversion of the scattering intensity  $I(q)$  using GNOM<sup>62</sup>.

**Ab initio shape determination.** *Ab initio* protein shapes were calculated from the experimental SAXS data using the bead-modeling program DAMMIN<sup>30</sup>. At least twenty independent fits were carried out without imposition of symmetry for the monomeric VirFG<sup>NDD</sup>, for the VirA<sup>CDD</sup>/VirFG<sup>NDD</sup> and VirA ACP<sub>5b</sub>-<sup>CDD</sup>/VirFG<sup>NDD</sup> non-covalent complexes, and the VirA<sup>CDD</sup>-VirFG<sup>NDD</sup> and VirA ACP<sub>5b</sub>-<sup>CDD</sup>-VirFG<sup>NDD</sup> covalent fusions, and with a 2-fold symmetry restraint for the VirA ACP<sub>5b</sub>-<sup>CDD</sup>-VirFG<sup>NDD</sup>-KS<sub>6</sub> fusion protein (in the absence of such an imposed symmetry, the fit between the experimental and model SAXS curves was poorer (greater  $\chi^2$ )). The results of several DAMMIN models were then compared using the program SUPCOMB<sup>47</sup> in order to determine a consensus model. The quality of the 3D modeling was determined using the discrepancy  $\chi^2$ , defined according to Konarev *et al.*<sup>61</sup> Low resolution models were then averaged using the DAMAVER and DAMFILT packages<sup>61,63</sup>. High-resolution structures or models were then superimposed into the *ab initio* protein shapes using SUPCOMB<sup>47</sup>. Figures displaying the protein structures were generated using PYMOL (Schrödinger, LLC) and the low-resolution bead models from DAMMIN were converted into mesh.

**Molecular modeling.** A three-dimensional structural model of the KS<sub>6</sub>-linker region was generated by pair-wise comparison of profile hidden Markov models using the HHpred server (<http://toolkit.tuebingen.mpg.de/hhpred>)<sup>64</sup>. Via this program, local sequence alignment using PSI-BLAST was performed and a multiple sequence alignment including information about predicted secondary structures was produced. The closest homologue to the VirA KS<sub>6</sub>-linker within the Protein Data Bank is the PksJ KS (PDB: 2NA1)<sup>65</sup>. This structure was then used to construct a model of the KS<sub>6</sub>-linker using the associated MODELLER software<sup>66</sup>. The KS<sub>6</sub> homodimer was reconstituted by superimposing two models onto the dimeric crystal structures of PksJ KS.

**Demonstration by SAXS that the covalent fusion of VirA<sup>CDD</sup> and VirFG<sup>NDD</sup> does not alter their interaction.** In order to provide experimental evidence that covalently fusing VirA<sup>CDD</sup> and VirFG<sup>NDD</sup> did not affect the mode of interaction between the two domains, we collected SAXS data on both VirA<sup>CDD</sup>-VirFG<sup>NDD</sup> fusion and the non-covalent complex VirA<sup>CDD</sup>/VirFG<sup>NDD</sup>. The theoretical scattering curve calculated from the NMR structure of the fusion VirA<sup>CDD</sup>-VirFG<sup>NDD</sup> was then compared with both the SAXS data obtained on the VirA<sup>CDD</sup>-VirFG<sup>NDD</sup> fusion and on the non-covalent complex VirA<sup>CDD</sup>/VirFG<sup>NDD</sup> using the program CRY SOL<sup>41</sup>, yielding good fits to the experimental data for scattering angles between 0.0005 and 0.4 Å<sup>-1</sup>, with comparable  $\chi^2$  of 2.61 and 2.93, respectively.

**Demonstration by SAXS that VirFG<sup>NDD</sup> undergoes a conformational change.** *Ab initio* molecular forms for VirFG<sup>NDD</sup> were calculated using DAMMIN<sup>30</sup>. 15 of the resulting low-resolution models ( $\chi^2$  of 2.32 vs. the SAXS data) were compared using SUPCOMB<sup>47</sup>, revealing an extended shape that is not consistent with the conformation of VirFG<sup>NDD</sup> adopted in the VirA<sup>CDD</sup>-VirFG<sup>NDD</sup> fusion. Indeed, calculation of a theoretical scattering curve for VirFG<sup>NDD</sup> alone using CRY SOL<sup>41</sup> on the basis of its structure within the VirA<sup>CDD</sup>-VirFG<sup>NDD</sup> fusion gives a poor fit to its experimental data for scattering angles between 0.0005 and 0.3 Å<sup>-1</sup> ( $\chi^2 = 31.55$ ). Furthermore, the Kratky plot<sup>31</sup> computed from the experimental SAXS curve is

consistent with a flexible particle (**Supplementary Fig. 8c**). Taken together, these calculations clearly show that free VirFG<sup>N</sup>DD and VirFG<sup>N</sup>DD in complex with partner do not adopt the same structure, consistent with an induced conformational change.

Supplementary Information

Novel docking interactions mediate intersubunit communication in the virginiamycin *trans*-acyl transferase polyketide synthase

Jonathan Dorival<sup>1,4</sup>, Thibault Annava<sup>1,4</sup>, Fanny Rissler<sup>1</sup>, Sabrina Collin<sup>1</sup>, Pierre Roblin<sup>2,3</sup>, Christophe Jacob<sup>1</sup>, Arnaud Gruetz<sup>1\*</sup>, Benjamin Chagot<sup>1\*</sup> and Kira J. Weissman<sup>1\*</sup>

<sup>1</sup>UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandœuvre-lès-Nancy Cedex, France. <sup>2</sup>SOLEIL Synchrotron, 91192 Gif-sur-Yvette, France. <sup>3</sup>INRA-URBIA, 44316 Nantes, France. \*These authors contributed equally to this work  
\*e-mail: arnaud.gruetz@univ-lorraine.fr, benjamin.chagot@univ-lorraine.fr or kira.weissman@univ-lorraine.fr

CONTENTS

Supplementary Results

Figure S1 | Comparative sequence analysis of VirFG<sup>DD</sup> and Class 2 docking domains.

Table S1 | Analysis of intersubunit interfaces in *trans*-AT PKSs.

Figure S2 | Identification of putative families of docking domains in *trans*-AT PKSs.

Table S2 | Putative docking domain families in *trans*-AT PKSs.

Figure S3 | SDS-PAGE analysis of protein preparations used in this study.

Table S3 | Sequences of constructs investigated in this study.

Figure S4 | Analysis by gel filtration of the interaction between VirA<sup>DD</sup> and VirFG<sup>DD</sup>.

Figure S5 | Determination by isothermal titration calorimetry of docking affinity.

Figure S6 | NMR studies of VirA<sup>DD</sup> and VirFG<sup>DD</sup>.

Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains.

Table S4 | Summary of SAXS data.

Figure S8 | Analysis by SAXS of the various constructs.

Table S5 | NMR and refinement statistics for VirA<sup>DD</sup>-VirFG<sup>DD</sup>.

Table S6 | Putative interface residues in the VirA<sup>DD</sup>/VirFG<sup>DD</sup> docking domain families.

Supplementary Methods

Table S7 | Primers used in this study.

Supplementary References

Supplementary Note

a  
MDAKEIITRFKDGQI<sup>28</sup>SRAAQALLAGR<sup>28</sup>DAAPRPSPAAAPTRPAVPAVPEFAAGTTVAE<sup>28</sup>GTAGRP

b  
CurL 1  
CurI 1  
CurJ 1  
CurG 1  
CurH 1  
CurK 1  
RhiC 1  
MmpA 1  
CurA 1  
Jame 1  
StiB 1  
TaiE 1  
Tai 1  
AjuA 1  
ElaK 1  
AjuB 1  
VirA 1  
SnaE2 1  
TaiL 1  
VirFG 1  
StiA 1  
StiG 1  
AjuJ 1  
AjuI 1  
consensus 1

CurL 3  
CurI 10  
CurJ 9  
CurG 13  
CurH 12  
CurK 12  
30  
1  
23  
24  
19  
39  
10  
AjuA 21  
AjuB 4  
AjuE 17  
15  
SnaE 5  
Tai 76  
StiA 21  
StiE 23  
StiG 10  
AjuJ 16  
AjuI 17  
consensus 81

c  
CurL 1  
CurH 1  
AjuF 1  
StiF 1  
StiB 1  
CurJ 1  
CurK 1  
CurM 1  
StiH 1  
AjuB 1  
CurI 1  
AjuC 1  
consensus 1

Figure S1 | Comparative sequence analysis of VirFG<sup>DD</sup> and Class 2 docking domains. (a) Sequence of subunit VirFG upstream of the conserved KS domain. The first 28 residues (in red) have been designated as VirFG<sup>DD</sup>. (b) Multiple alignment of the C-terminal regions of PKS subunits from *trans*-AT PKSs (in red), with those of *cis*-AT PKS<sup>1</sup> (black, except for those of subunit CurA and the closely-related Jame (in blue)) (analysis as published in ref. 2). The sequences (with the exception of VirFG<sup>DD</sup>) from the *trans*-AT PKSs lie downstream of the conserved sequences of single (Tai1) or tandem acyl carrier protein (ACP) domains associated with  $\beta$ -modification reactions, which are located at the ends of subunits. The *trans*-AT PKS regions were identified manually, as BLAST analysis using the VirA region did not yield any homologous sequences. The positions of the two  $\alpha$ -helices in the solved structure of the CurG and CurK C-terminal docking domains are indicated, as are the hydrophobic (H) and charged (C) residues which contribute to the docking interface<sup>1</sup>. (c) Multiple alignment of the N-terminal regions from the partner *cis*-AT PKS subunits to those shown in b, alongside that of VirFG (in red). The positions of the two  $\alpha$ -helices in the solved structure of the CurH and CurL N-terminal docking domains are indicated, as are the hydrophobic (H) and charged (C) residues which contribute to the docking interface<sup>1</sup>. Key: Cur, curacin (*cis*-AT PKS)<sup>2</sup>; Rhi, rhizoxin (*trans*-AT)<sup>3</sup>; Mmp, mupirocin (*trans*-AT)<sup>4</sup>; Jame, jamaicamide (*cis*-AT)<sup>5</sup>; Sti, stigmatellin (*cis*-AT)<sup>7</sup>; Tai, thailandamide (*trans*-AT)<sup>8</sup>; Ta, myxovirescin (antibiotic Ta; *trans*-AT)<sup>9</sup>; Aju, ajudazol (*cis*-AT)<sup>10</sup>; Ela, elansolid (*trans*-AT)<sup>11</sup>; Vir, virginiamycin (*trans*-AT)<sup>12</sup>; Sna, pristinamycin (*trans*-AT)<sup>13</sup>.

Table S1 | Analysis of intersubunit interfaces in *trans*-AT PKSs.

C-terminal domain	N-terminal domain	Split module?	Systems where present (number of interfaces if >1)
ACP	KS		Chi (2), Cor, Ela, FR9, Kir, Mmp, Ooc, Ozm (3), Rhi (2), Sna, Tai (2), Tdp, Tmp (2), Vir, 592 (5)
ACP	KS*		Cal (3), Mmp, Nsp, Onn, Ped, Psy
ACP	DH	✓	Cor
ACP	DH*	✓	Bry
ACP	ER	✓	Bry (2), Etn, Ta
ACP	C		Bae, Bat, Chi, Kir (2), Ozm, Sna (2), Tai, Tdp, Vir*
ACP	HC		Dis, Riz, Vir
KS	DH	✓	Chi, Dif (2), Ela, Etn (2), FR9(2), Kir (2), Lnm, Mln (2), Sor (3), Tst
KS*	DH	✓	Bae (2), Bat, Bon, Cal, Dif, Rhi, Tai
KS	DH*	✓	Cor, Dif
KS	C-MT	✓	Ta, Tai
KS	KR	✓	Ela, Lac, Lkc (2), Mgs, Mln, Riz
KS	KR*	✓	592
KS	KS	✓	Ooc
KS	ECH	✓	Cal
FH	ACP	✓	Bae, Bon, Cal, Dif (2), Ela, Etn, Ooc, Rhi, Sor (3)
KR	C-MT	✓	Bon, Etn
KR	ACP	✓	Lkc
DH	ACP	✓	Lac, Mgs, Ped
DH	KR	✓	Dis, Lkc (2), Mln (2), Tst
DH*	KR	✓	Mln
PS	KR	✓	Sor
C	A	✓	Alb
HC	A	✓	Cal, Rhi
C	GNAT	✓	Bry
C	C	✓	FR9, Tst
C	N-MT	✓	Ttt
PCP	KS	✓	Ozm, Nsp, Sna, Tdp(2)
O-MT	ACP	✓	Ozm
C-MT	ACP	✓	FR9, Riz
Ox	KR	✓	FR9, Tst
E	C	✓	Tdp

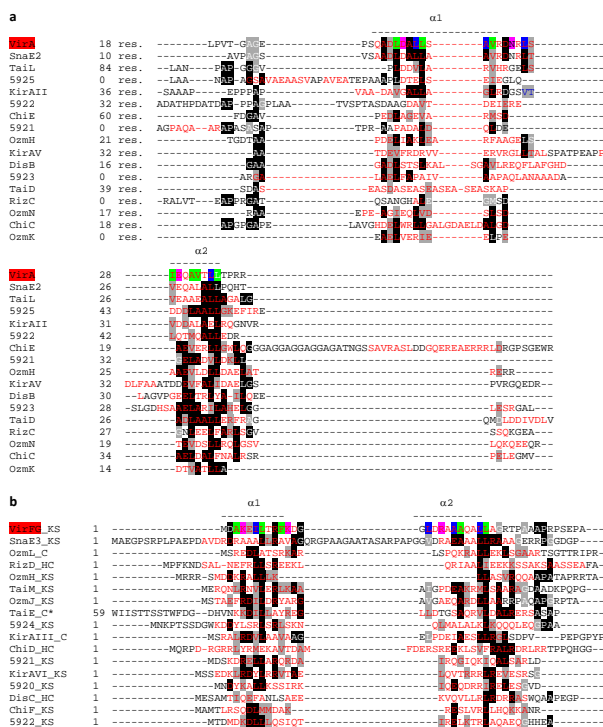
\*The full Vir-encoding region has not been published<sup>12</sup>. But based on comparison with the highly-homologous *sno* cluster from *S. pristinaespiralis* which gives rise to the same molecule<sup>43</sup>, we presume that the Vir PKS also incorporates such an interface.

The first two columns indicate the functional domains present at the C- and N-termini, respectively, of the interacting subunits, while the third reports whether the interface constitutes a split module.

**Key to domain names:** KS, ketosynthase; DH, dehydratase; C-MT, C-methyltransferase; KR, ketoreductase; FH, FkbH-like domain; ACP, acyl carrier protein; C, condensation; HC, heterocyclization; ER, enoyl reductase; PS, pyrane synthase; A, adenylation; GNAT, Gcn5-related N-acetyltransferase; PCP, peptidyl carrier protein; O-MT, O-methyltransferase; Ox, oxidase; E, epimerase; \*, inactive.

**Key to *trans*-AT PKS systems:**

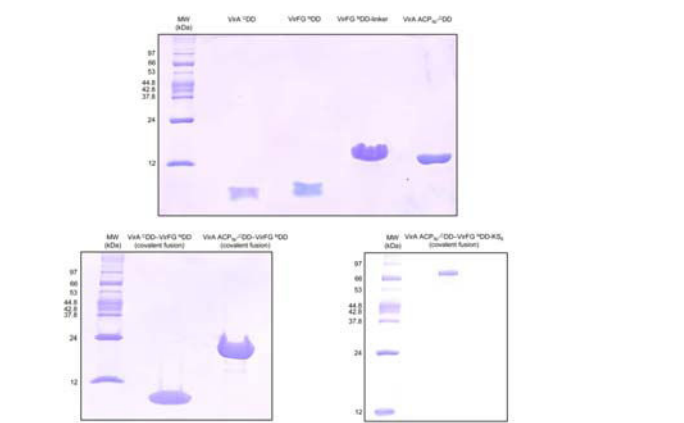
Chi, chivosazol<sup>14</sup>; Cor, corallopyronin<sup>15</sup>; Ela, elansolid<sup>11</sup>; FR9, FR901464<sup>16</sup>; Kir, kirromycin<sup>17</sup>; Mmp, mupirocin<sup>4</sup>; Ooc, oocydin<sup>18</sup>; Ozm, oxazolomycin<sup>19</sup>; Rhi, rhizoxin<sup>3</sup>; Sna, pristinamycin<sup>13</sup>; Tai, thailandamide<sup>8</sup>; Tdp, thailandepsin<sup>20</sup>; Tmp, thiomarinol<sup>21</sup>; Vir, virginiamycin<sup>12</sup>; 592, encyloxin<sup>22</sup>; Cal, calyculin<sup>23</sup>; Nsp, nosperin<sup>24</sup>; Onn, onnamide<sup>25</sup>; Ped, pederin<sup>26</sup>; Psy, psymberin<sup>27</sup>; Bry, bryostatin<sup>28</sup>; Etn, etnangien<sup>29</sup>; Ta, myxovirescin (antibiotics TA)<sup>9</sup>; Bae, bacillaene<sup>30</sup>; Bat, kalimantarin<sup>31</sup>; Dis, disorazol<sup>32,33</sup>; Riz, rhizopodin<sup>34</sup>; Dif, diffidicin<sup>35</sup>; Lnm, leinamycin<sup>36</sup>; Mln, macrolactin<sup>37</sup>; Sor, sorangicin<sup>38</sup>; Tst, thalinstatin<sup>39</sup>; Bon, bongkrekic acid<sup>40</sup>; Lac, lactidomycin<sup>41</sup>; Lkc, lankacidin<sup>42</sup>; Mgs, migrastatin<sup>43</sup>; Alb, albidicin<sup>44</sup>; Ttt, thalassospiramide<sup>45</sup>.



**Figure S2 | Identification of putative families of docking domains in *trans*-AT PKSs.** (a) Secondary structure-based alignment by PsiPred<sup>®</sup> of VirA<sup>C</sup>DD alongside putative C-terminal docking domains from *trans*-AT PKSs operating at junctions between ACP various partner domains. Predicted  $\alpha$ -helices are indicated in red, and the extents of helices  $\alpha 1$  and  $\alpha 2$  in VirA<sup>C</sup>DD as determined by NMR are shown above the alignments. For clarity, a number of amino acids between the upstream conserved ACP domain and the displayed sequence have been omitted, as indicated (e.g. 18 res. for VirA<sup>C</sup>DD). (b) Secondary structure-based alignment of VirFG<sup>N</sup>DD alongside the putative partner docking domains of those shown in a. In all cases, the downstream catalytic domain is indicated: KS, ketosynthase; C, condensation; HC, heterocyclization). Key to systems: Vir, virginiamycin; Sna, pristinamycin; Tai, thailandamide; S92, enacyloxin; Kir, kirromycin; Chi, chivosazol; Ozm, oxazolomycin; Dis, disorazol; Riz, rhizopodin. Key to color coding: **green** = residue at hydrophobic interface in VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD complex structure; **blue** = intradomain hydrophobic interaction residue; **pink** = residue participating in an electrostatic interaction at the docking interface. \*This N-terminal region incorporates an additional 60 residues (not shown).

Table S2   Putative docking domain families in <i>trans</i> -AT PKSs.			
	C-terminal domain	N-terminal domain	Systems/interfaces where present <sup>a</sup>
I	ACP	KS	ChiE/ChiF; KirAV/KirAVI; OzmH/J and OzmN/H; SnaE2/SnaE3; TaiL (also ACP-ACP)/TaiM; VirA/VirFG; S921/S920, S922/S921, S923/S922 and S925/S924
	ACP	C/HC	ChiC/ChiJ; DisB/DisC; RizC/RizD; KirAll/KirAllI; OzmK/OzmL; TaiD/TaiE
II	ACP	KS	Elak/Elao; RhiC/RhiD; TdpB/TdpC1
	ACP	ER	BryA/BryB and BryB/BryC; EtnH/EtnI
	ACP	C	BaeM/BaeN; Bat1/Bat2
III	ACP	KS	ChiB/ChiC; S920/S919; TaiE/TaiK
	ACP	C	TdpC1/TdpD1

<sup>a</sup>The host organisms of the seven family I systems include two *Streptomyces* strains (*Streptomyces collinus* Tu 365 (kirromycin) and *Streptomyces albus* IA43453 (oxazolomycin)), three myxobacterial strains (*Sorangium cellulosum* Soc 56 (chivosazol), *Sorangium cellulosum* Soc 12 (disorazol) and *Stigmatella aurantiaca* Sg a15 (rhizopodin)) and two *Burkholderia* strains (*Burkholderia thailandensis* (thailandamide), and *Burkholderia ambifaria* AMMD (enacyloxin)). Thus, these docking domains show a limited phylogenetic distribution, but nonetheless occur within multiple phyla (actino- and proteobacteria). Interfaces derived from the same system are indicated by matching colors (i.e. red for enacyloxin).



**Figure S3 | SDS-PAGE analysis of protein preparations used in this study.** VirA<sup>C</sup>DD (calc'd: 4.3 kDa); VirFG<sup>N</sup>DD (calc'd: 4.1 kDa); VirFG<sup>N</sup>DD-linker (calc'd: 7.2 kDa); VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD (covalent fusion) (calc'd: 8.4 kDa); VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD-KS (covalent fusion) (calc'd: 85.4 kDa). The molecular weights of the markers are indicated.

Table S3   Sequences of constructs investigated in this study.	
Name of construct	Sequence <sup>a</sup>
VirA <sup>C</sup> DD	GGGSVYTGAGEPQQADLDALLSAVRDNRLSIEQAVTLLTPRR
VirFG <sup>N</sup> DD	GGGSMDAKEILTRFKDGLDRRAAQALLAGRTPAAPRPY
VirFG <sup>N</sup> DD-linker	GGGSMDAKEILTRFKDGLDRRAAQALLAGRTPAAPRPSEPAAPPTPVAAPVAVPEPAA GTTVAAGTGRGP
VirA <sup>C</sup> DD-VirFG <sup>N</sup> DD	GGGSVAVDPAVARALREELARTLYCEPGDIDDEASFNLLGLDLSLGVFVAVFNQTYGL DEKAGILLYDHSALALSRHVAGRAAPVAAAAPALPLPV
VirFG <sup>N</sup> DD-KS <sub>6</sub>	GGGSMDAKEILTRFKDGLDRRAAQALLAGRTPAAPRPSEPAAPPTPVAAPVAVPEPAA GTTVAAGTGRGPVAVVIGYARFPGAADADTFWQRILDDGDLVTEVPPERWRTEFYDP DPAABGRSVSRWGAVVADADRFDADFRFMPREABLTDQARLFLQEAARALEHAGRDARS LAGTRCGVYAGVMLNDYQDLVERESPYPKRLPQVMQNSNSILAARIAVHLDLGGPVTVD ACSSSLTALLACQSLWGETLIVVGVVTLTFLPHVMSAGMLSPNGRCRPFDAAD GIVPGCCAVVTLKPLSKALADGDPVAVIRASGLNDGPGQGGITAPASRQATLVRTLL QRFPVDPAGIDVVECHGTGTPLDGDIETVALNEAFAGLADPASVPIGSKMIGHTSAA AGLAGLKAAGVVRVGLVPPSLHYARANPQIPFDQGPFTVAGERRBLGRPDGRRPRATVS SPGFSGTNAVYVVEQAPQARPAAGDDGAPVLPVPLSGPARADAPAAHADLARLWRGPGS DASPADVHTLAVARTHHTYRSALLVSGRDELLESLLDLAGSADPRRTDAPDAAPDAA VRRQAALLLARLVERAGENPGVELAALAKLYTQHTVDAAVSPPARHRLSLPAYFPAP HRHIVER
VirA <sup>C</sup> DD-VirFG <sup>N</sup> DD	GGGSVYTGAGEPQQADLDALLSAVRDNRLSIEQAVTLLTPRRGGSGGGSMDAKEILTRFKD GGLDRAAQALLAGRTPAAPRP
VirA <sup>C</sup> DD-VirFG <sup>N</sup> DD	GGGSVAVDPAVARALREELARTLYCEPGDIDDEASFNLLGLDLSLGVFVAVFNQTYGL DEKAGILLYDHSALALSRHVAGRAAPVAAAAPALPLPVTTGAGEPQQADLDALLSAVRD NRLSIEQAVTLLTPRRGGSGGGSMDAKEILTRFKDGLDRRAAQALLAGRTPAAPRP NRLSIEQAVTLLTPRRGGSGGGSMDAKEILTRFKDGLDRRAAQALLAGRTPAAPRPSE PAAAPTTPVAAPVAVPEPAAAGTVAAGTGRGPVAVVIGYARFPGAADADTFWQRILDD GDLVTEVPPERWRTEFYDPAABGRSVSRWGAVVADADRFDADFRFMPREABLTDQAR LFLQEAARALEHAGRDARSLAGTRCGVYAGVMLNDYQDLVERESPYPKRLPQVMQNSNSI LAARIAVHLDLGGPVTVDACSSSLTALLACQSLWGETLIVVGVVTLTFLPHVMSAGMLSP NAGMLSPNGRCRPFDAADGIVPGCCAVVTLKPLSKALADGDPVAVIRASGLNDGPGQGGI TAPASRQATLVRTLLQRFPVDPAGIDVVECHGTGTPLDGDIETVALNEAFAGLADPASV PIGSKMIGHTSAAAGLAGLKAAGVVRVGLVPPSLHYARANPQIPFDQGPFTVAGERRBLGR PDGRRPRATVSPPGFSGTNAVYVVEQAPQARPAAGDDGAPVLPVPLSGPARADAPAAHAD LARLWRGPGSDASPADVHTLAVARTHHTYRSALLVSGRDELLESLLDLAGSADPRRTDAP DAAPDAAVRRQAALLLARLVERAGENPGVELAALAKLYTQHTVDAAVSPPARHRLSLPAY FPAPHRHIVER
VirA <sup>C</sup> DD-VirFG <sup>N</sup> DD-KS <sub>6</sub>	GGGSVYTGAGEPQQADLDALLSAVRDNRLSIEQAVTLLTPRRGGSGGGSMDAKEILTRFKD GGLDRAAQALLAGRTPAAPRP

<sup>a</sup>Red = sequence introduced from the plasmid; green = added Tyr; blue = linker used to fuse the docking domains.

a

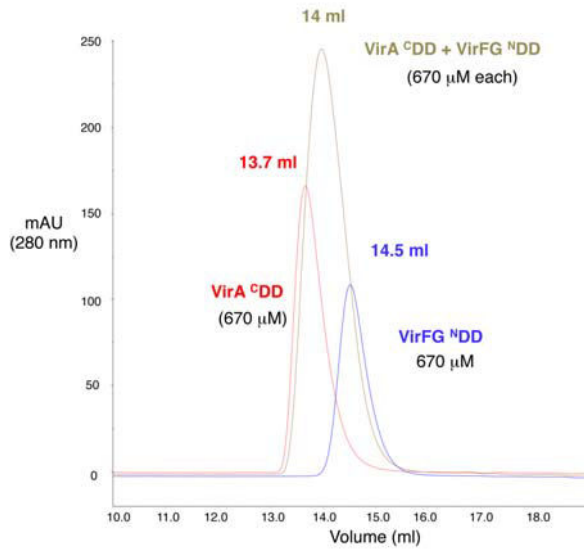


Figure S4 | Analysis by gel filtration of the interaction between VirA<sup>Δ</sup>DD and VirFG<sup>Δ</sup>DD. The individual docking domains VirA<sup>Δ</sup>DD and VirFG<sup>Δ</sup>DD (both at an initial concentration of 670 μM) exhibit elution volumes of 13.7 ml and 14.5 ml, respectively, leading to MW estimates of 11.8 and 7.4 kDa. When combined, a new peak is observed at an intermediate elution volume (14 ml), translating into an estimated molecular weight of 9.9 kDa.

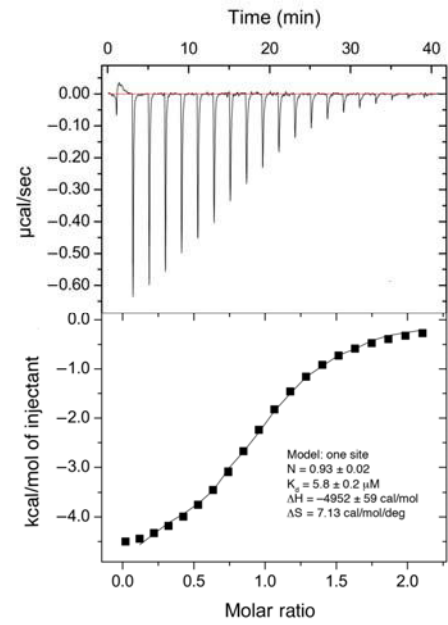


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (a) Analysis of binding between wild type VirA<sup>Δ</sup>DD and VirFG<sup>Δ</sup>DD. (One of three measurements is shown, but the  $K_d$  value represents the average of three independent experiments).

b

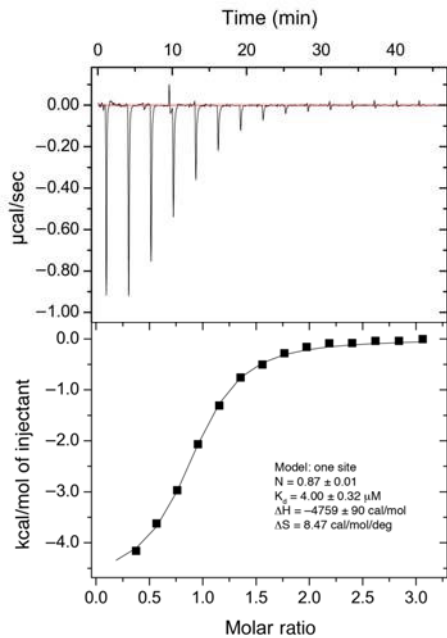


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (b) Analysis of binding between wild type VirA ACP<sub>30</sub>-<sup>Δ</sup>DD and VirFG<sup>Δ</sup>DD.

c

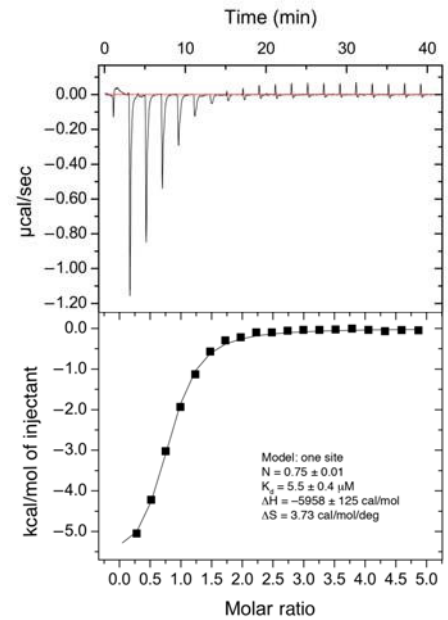


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (c) Analysis of binding between wild type VirA<sup>Δ</sup>DD and VirFG<sup>Δ</sup>DD-linker.

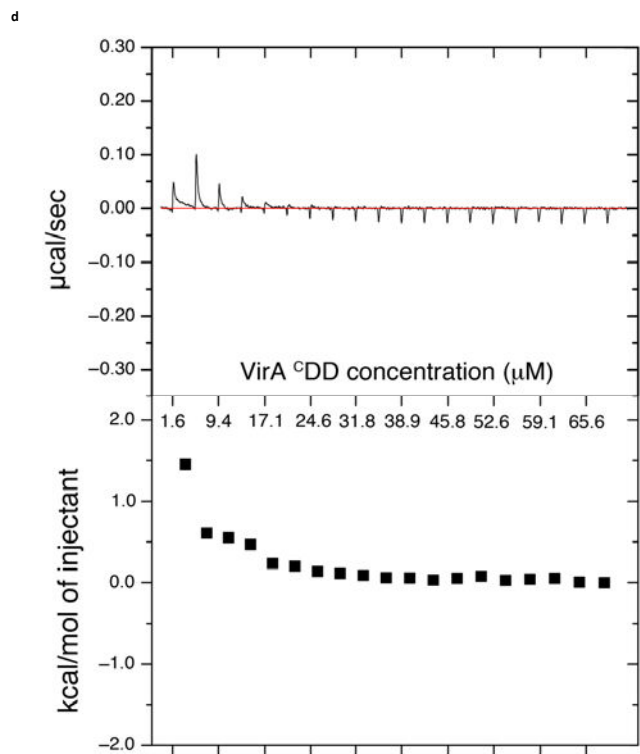


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (d) Dilution analysis of VirA<sup>Δ</sup>DD. No further change in heat signal is observed after VirA<sup>Δ</sup>DD reaches a concentration of 30 μM, indicating that it is completely homodimeric above this concentration.

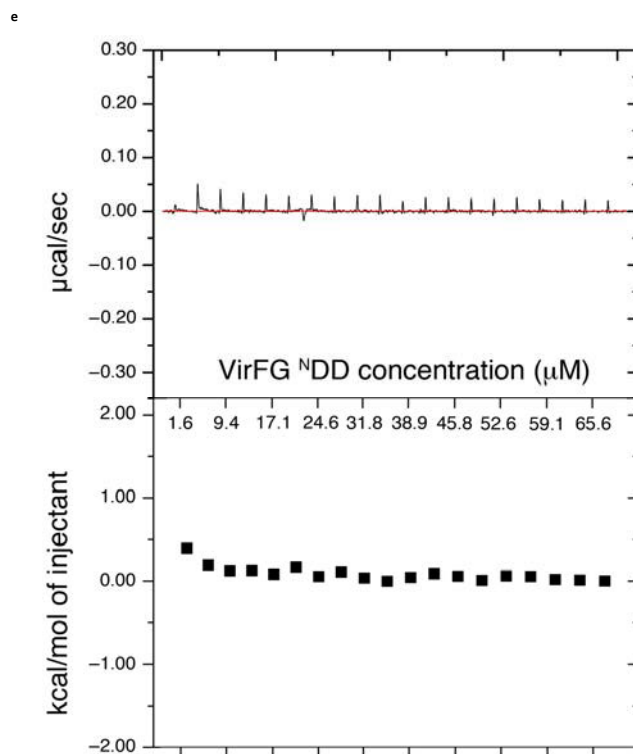


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (e) Dilution analysis of VirFG<sup>N</sup>DD. No evidence for homodimerization is observed across the full range of concentrations (to 78 μM).

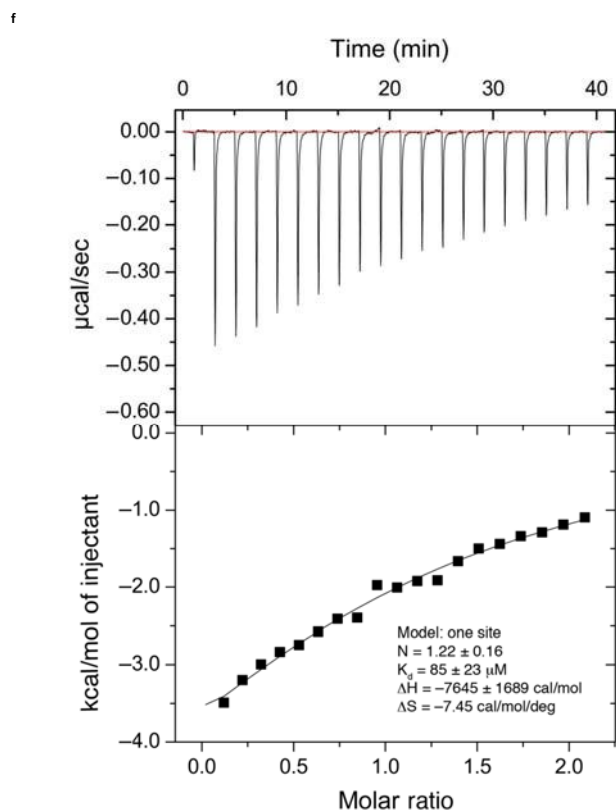


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (f) Analysis of binding between VirA<sup>Δ</sup>DD D6940A and wild type VirFG<sup>N</sup>DD.

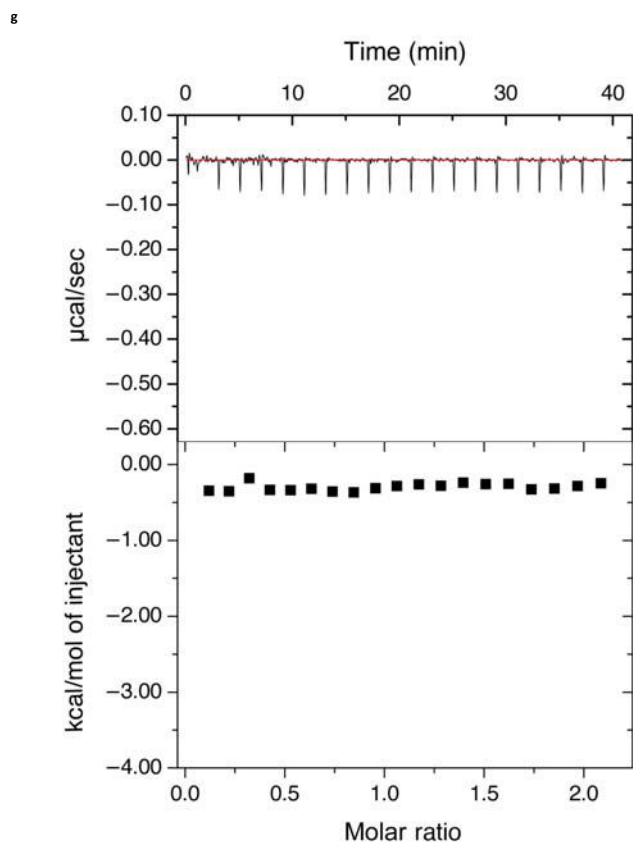


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (g) Analysis of binding between VirA<sup>Δ</sup>DD D6940R and wild type VirFG<sup>N</sup>DD.



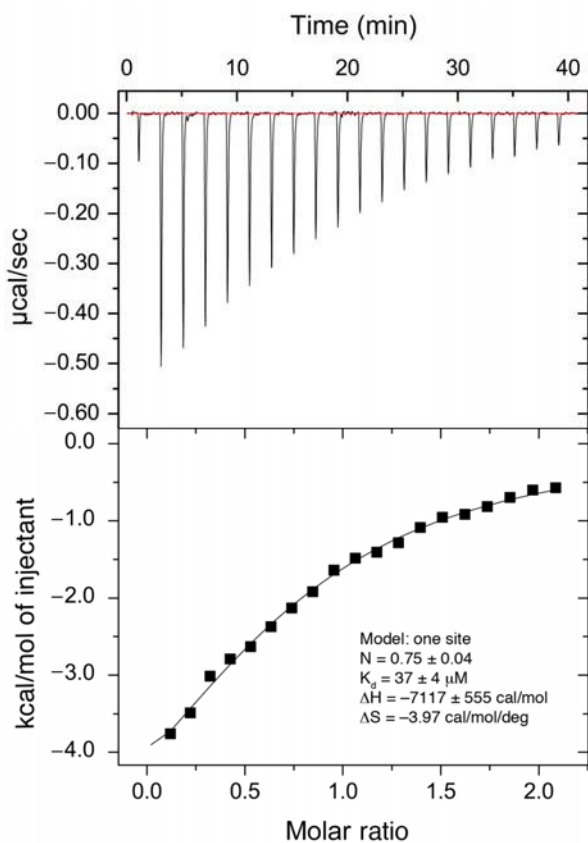


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (h) Analysis of binding between VirA<sup>Δ</sup>DD N6949K and wild type VirFG<sup>Δ</sup>DD.

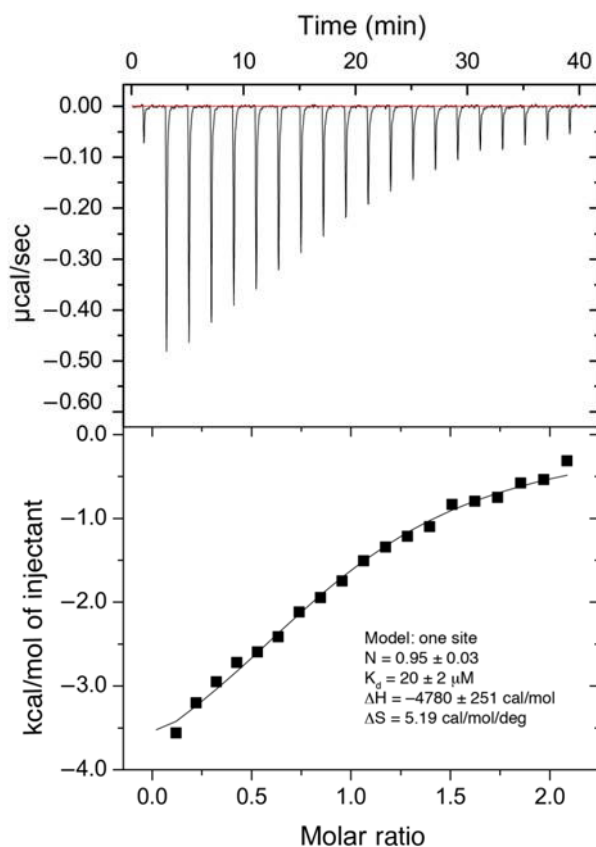


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (i) Analysis of binding between VirA<sup>Δ</sup>DD E6954A and wild type VirFG<sup>Δ</sup>DD.

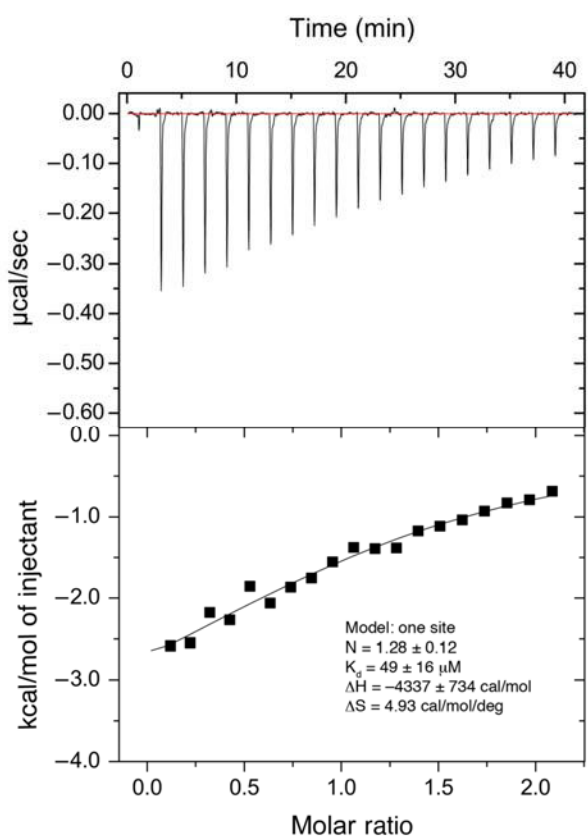


Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (j) Analysis of binding between VirA<sup>Δ</sup>DD E6954K and wild type VirFG<sup>Δ</sup>DD.

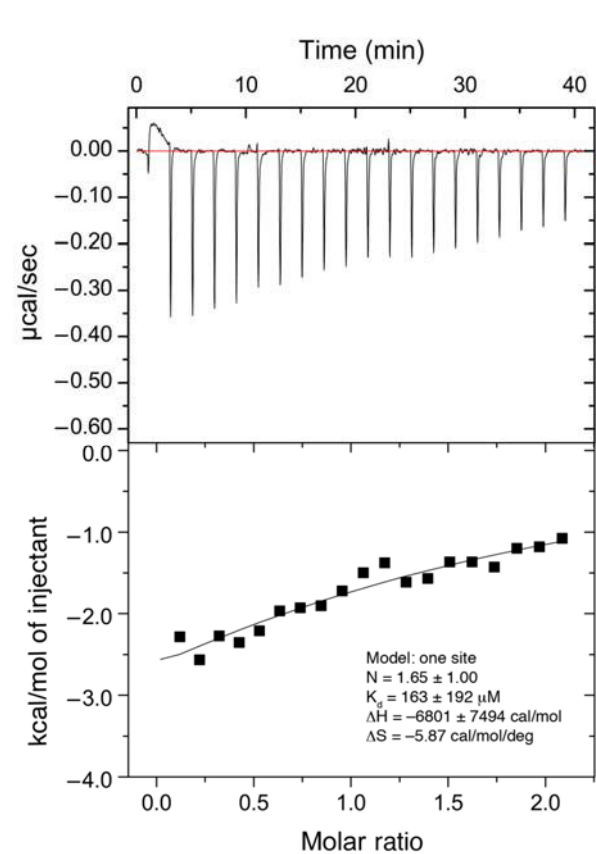
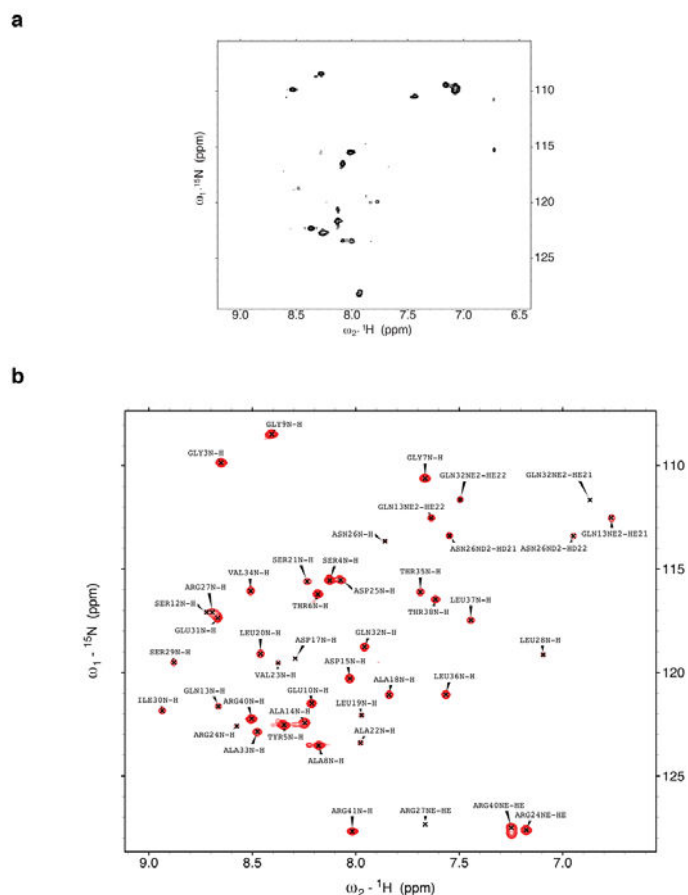
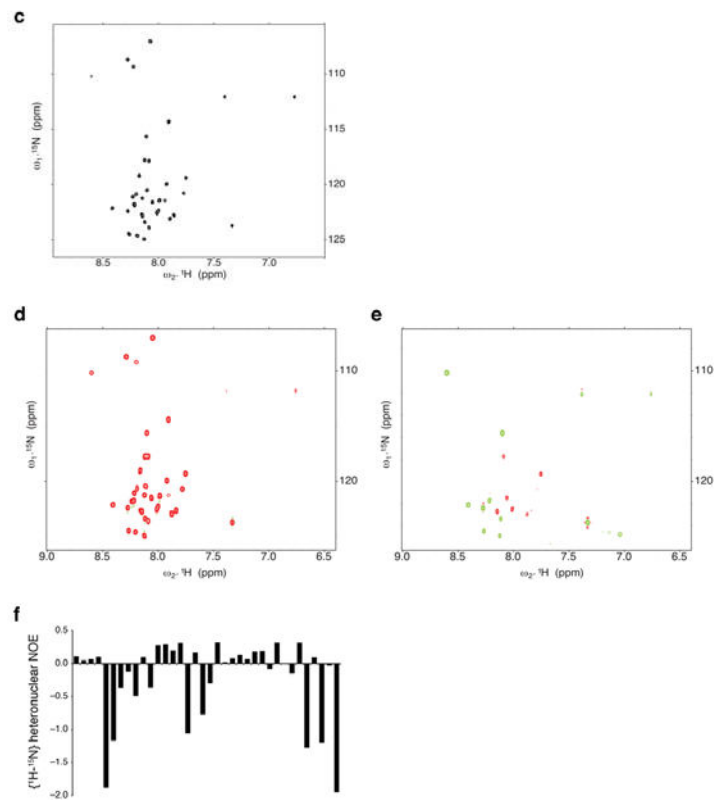


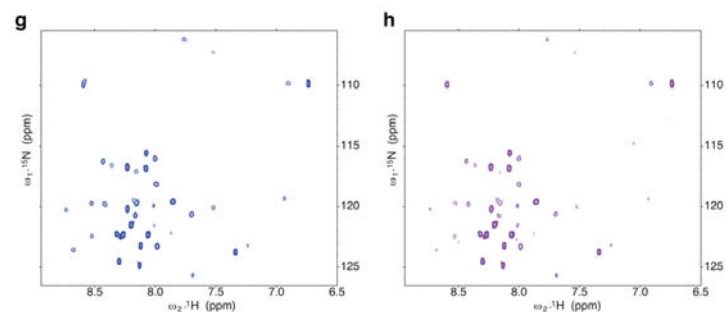
Figure S5 | Determination by isothermal titration calorimetry of docking affinity. (k) Analysis of binding between VirA<sup>Δ</sup>DD D6940A + E6954A and wild type VirFG<sup>Δ</sup>DD. For b-k, the indicated error is that on the individual measurement.



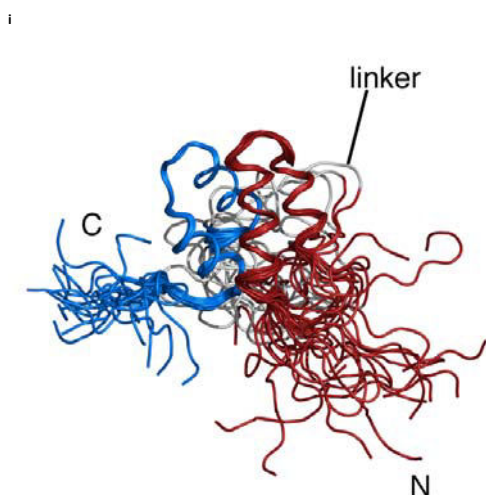
**Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (a) Unassigned [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum of <sup>15</sup>N VirA<sup>C</sup>DD alone. (b) Assigned [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum of <sup>15</sup>N VirA<sup>C</sup>DD in complex with <sup>15</sup>N VirFG<sup>N</sup>DD. Residues are labeled according to the construct numbers (Fig. 3a).



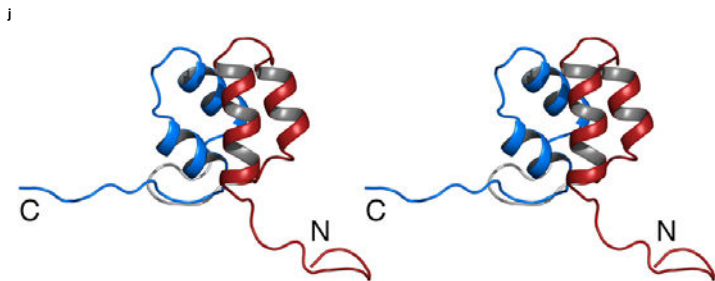
**Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (c) Unassigned [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum of <sup>15</sup>N VirFG<sup>N</sup>DD alone. (d,e,f) Dynamical studies of VirFG<sup>N</sup>DD. (d) Reference <sup>1</sup>H-<sup>15</sup>N heteronuclear NOE spectrum acquired at 600 MHz. (e) Saturated <sup>1</sup>H-<sup>15</sup>N heteronuclear NOE spectrum acquired at 600 MHz. (f) Resulting <sup>1</sup>H-<sup>15</sup>N heteronuclear NOE values plotted again unassigned peaks.



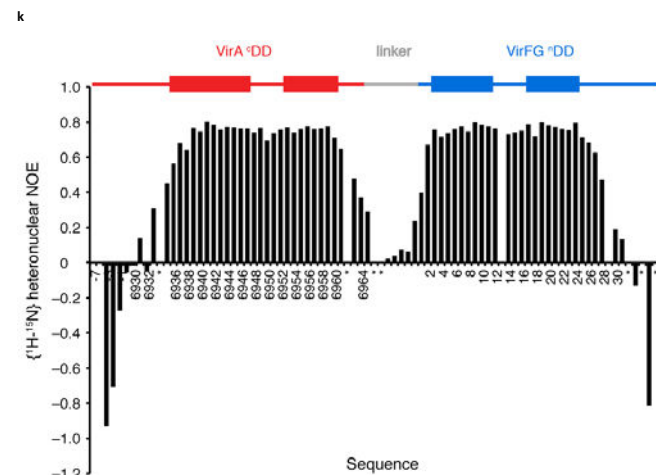
**Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (g) Unassigned [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum of VirFG<sup>N</sup>DD in complex with VirA<sup>C</sup>DD. (h) Unassigned [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectrum of VirFG<sup>N</sup>DD in complex with VirA<sup>C</sup>ACP<sub>30</sub>-DD.



**Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (i) The final ensemble of 20 conformers that represents the solution structure of VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD. VirA<sup>C</sup>DD is colored red and VirFG<sup>N</sup>DD blue. The introduced linker, N at C-termini are indicated.



**Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (j) Stereo-view of a representative VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD structure depicted as cartoon. VirA<sup>C</sup>DD is colored red and VirFG<sup>N</sup>DD blue. The introduced linker, N at C-termini are indicated.



**Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD.** (k) Dynamical study of VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD construct. <sup>1</sup>H-<sup>15</sup>N heteronuclear NOE values are plotted against the sequence. The construct regions are colored as followed: red for VirA<sup>C</sup>DD, gray for the linker and blue for VirFG<sup>N</sup>DD. The N and C-termini are highly flexible as evidenced by negative <sup>1</sup>H-<sup>15</sup>N heteronuclear NOE. The linker region for which <sup>1</sup>H-<sup>15</sup>N heteronuclear NOEs are lower than 0.5 is also very dynamic. \* = not possible to attribute the amide signal.

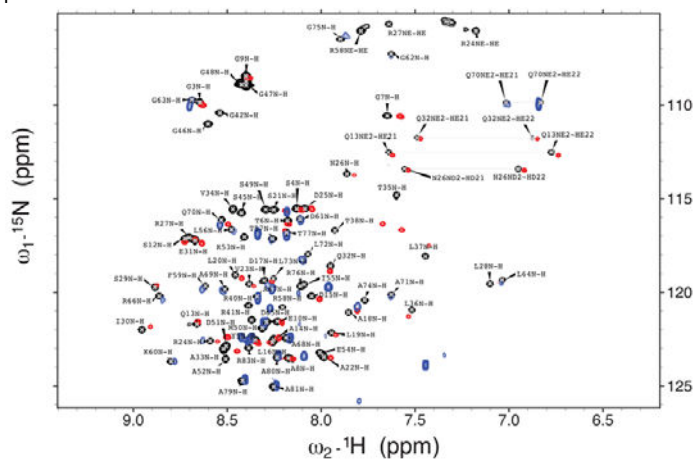


Figure S6 | NMR studies of VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD. (I) Superposition of the [<sup>1</sup>H,<sup>15</sup>N]-HSQC spectra of VirA<sup>C</sup>DD in complex with VirFG<sup>N</sup>DD (red), VirFG<sup>N</sup>DD in complex with VirA<sup>C</sup>DD (blue), and assigned VirA<sup>C</sup>DD-VirFG<sup>N</sup>DD (black).

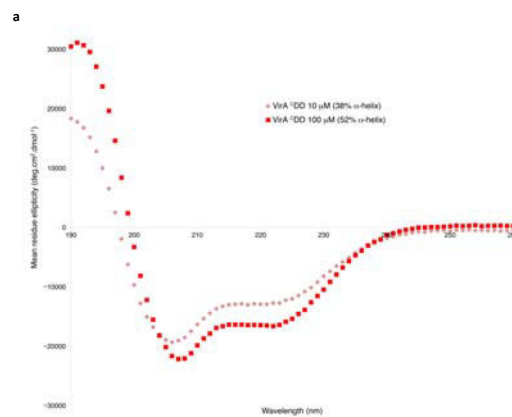


Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains. (a) Analysis of VirA<sup>C</sup>DD at concentrations above (100 μM) and below (10 μM) the point at which it homodimerizes completely. The percentage α-helix is indicated in each case.

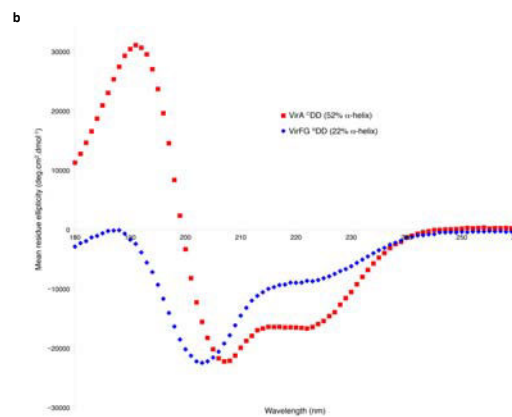


Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains. (b) Analysis of wild type VirFG<sup>N</sup>DD and for comparison, VirA<sup>C</sup>DD (both at 100 μM). The percentage α-helix is indicated in each case.

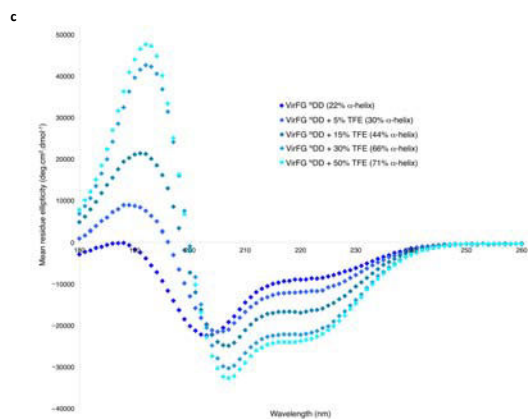


Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains. (c) Effect of various concentrations of 2,2,2-trifluoroethanol (TFE) on the structure of VirFG<sup>N</sup>DD. The percentage TFE and α-helix is indicated in each case.

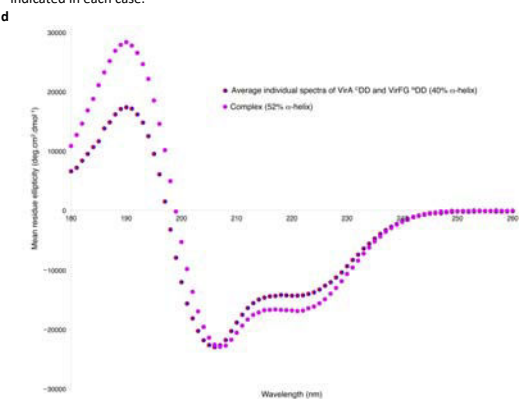


Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains. (d) Spectra showing net structuration upon combining VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD. A combined spectrum was obtained by analyzing simultaneously isolated VirA<sup>C</sup>DD and VirFG<sup>N</sup>DD (at 100 μM). For comparison, the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μM. The percentage α-helix is indicated in each case.

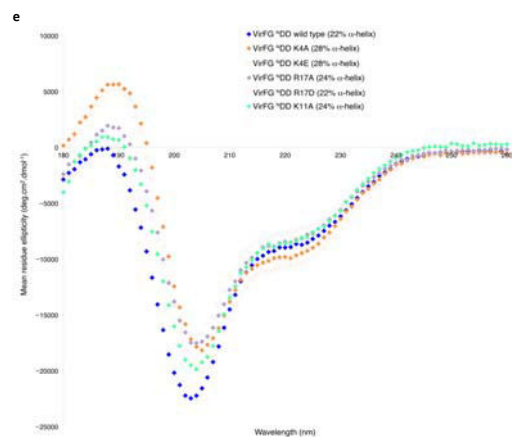


Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains. (e) Analysis of site-directed mutants of VirFG<sup>N</sup>DD. The introduced mutations resulted in substantial changes in secondary structure, as indicated by the calculated percentage α-helix.

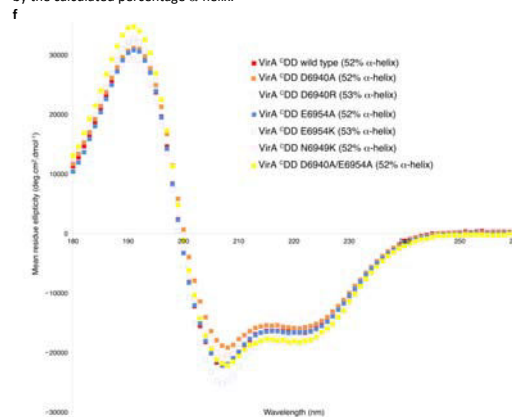


Figure S7 | Analysis by circular dichroism of wild type and mutant docking domains. (f) Analysis of site-directed mutants of VirA<sup>C</sup>DD. By and large, these mutations introduced minimal structural perturbation into the docking domain.

Construct	$R_g$ (Guinier) (Å) <sup>a</sup>	$R_g$ (GNOM) (Å) <sup>a</sup>	$D_{max}$ (GNOM) (Å) <sup>b</sup>	$\chi^2$ (DAMMIN) c	$\chi^2$ (CRYSOLO) c
VirFG <sup>h</sup> DD	16 ± 1	16.88 ± 0.01	66	2.32	31.55
VirA <sup>c</sup> DD/VirFG <sup>h</sup> DD complex	15.35 ± 0.01	15.76 ± 0.01	56	5.88	2.93
VirA <sup>c</sup> DD-VirFG <sup>h</sup> DD fusion	15.12 ± 0.01	15.49 ± 0.01	60	6.18	2.61
VirA ACP <sub>35</sub> - <sup>c</sup> DD/ VirFG <sup>h</sup> DD complex	28.3 ± 0.3	31.0 ± 0.2	120	1.29	n.d.
VirA ACP <sub>35</sub> - <sup>c</sup> DD <sup>2</sup>	37.0 ± 0.3	37.7	140	1.231	n.d.
VirA ACP <sub>35</sub> - <sup>c</sup> DD- <sup>h</sup> DD VirFG fusion	27 ± 2	29.49 ± 0.01	118	2.44	n.d.
VirA ACP <sub>35</sub> - <sup>c</sup> DD- <sup>h</sup> DD-KS <sub>5</sub> VirFG fusion	48 ± 2	50.5 ± 0.1	200	1.76	n.d.

<sup>a</sup> $R_g$  is the radius of gyration derived from the atomic models using the program PRIMUS<sup>47</sup>. For experimental data,  $R_g$  is given by the Guinier approximation.  
<sup>b</sup> $D_{max}$  is the maximum particle diameter derived from the distance distribution function  $P(r)$  using the program GNOM<sup>48</sup>.  
<sup>c</sup> $\chi^2$  is the discrepancy between the experimental SAXS profile and the theoretical curve calculated using the programs DAMMIN<sup>49</sup> or CRYSOLO<sup>50</sup>.  
 n.d.: not determined.

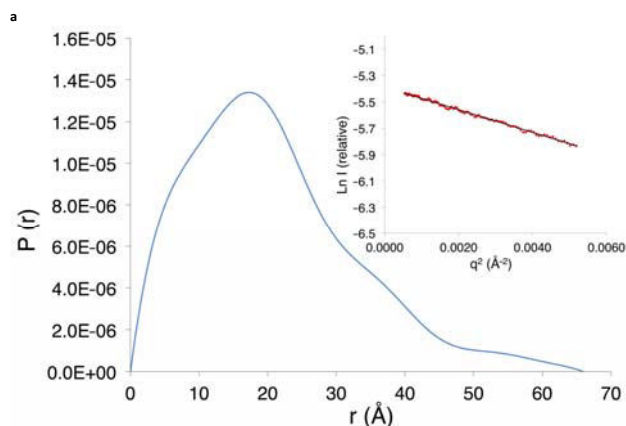


Figure S8 | Analysis by SAXS of the various constructs. (a) The distance distribution function for VirFG<sup>h</sup>DD calculated with GNOM. The Guinier plot is inset.

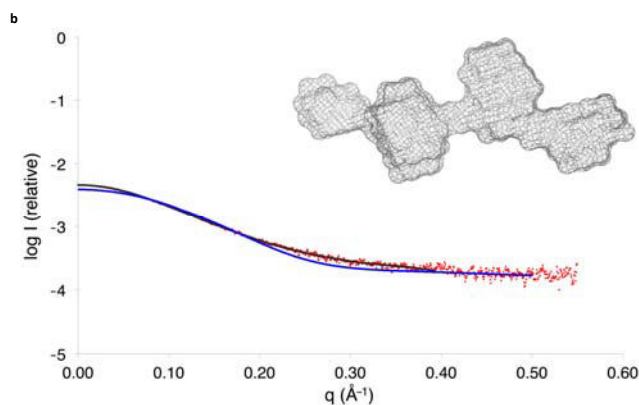


Figure S8 | Analysis by SAXS of the various constructs. (b) Representative superimposition of the experimental SAXS data acquired on VirFG<sup>h</sup>DD (red dots) and the theoretical curve for VirFG<sup>h</sup>DD calculated using CRYSOLO from the VirA<sup>c</sup>DD-VirFG<sup>h</sup>DD NMR structure (solid blue line) shows that the two diverge significantly. Fit between the *ab initio* model computed with DAMMIN (solid black line) with the experimental SAXS data (red dots) obtained on VirFG<sup>h</sup>DD. Inset (grey mesh), the *ab initio* envelope of VirFG<sup>h</sup>DD calculated from DAMMIN.

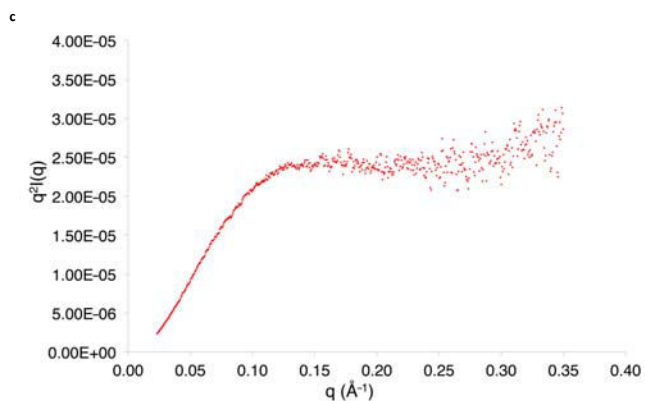


Figure S8 | Analysis by SAXS of the various constructs. (c) The Kratky plot (red dots) for VirFG<sup>h</sup>DD calculated in the  $q$  range of 0.0005–0.35 Å<sup>-1</sup>.

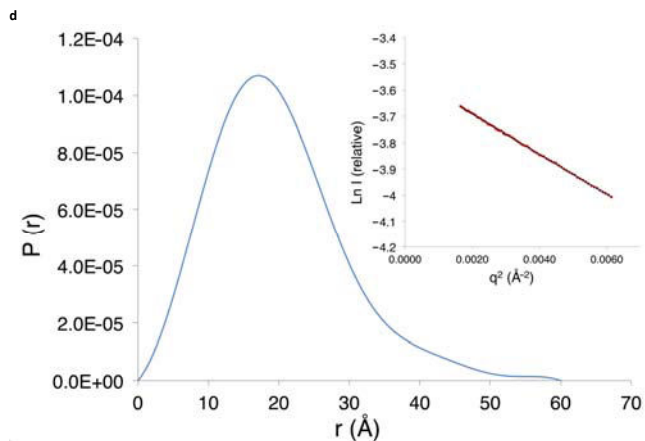


Figure S8 | Analysis by SAXS of the various constructs. (d) The distance distribution function for VirA<sup>c</sup>DD-VirFG<sup>h</sup>DD (covalent fusion) calculated with GNOM. The Guinier plot is inset.

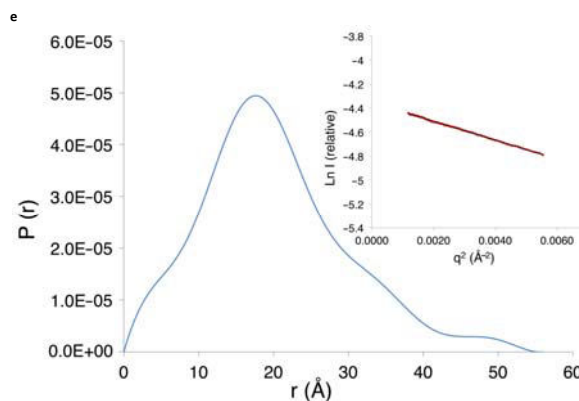


Figure S8 | Analysis by SAXS of the various constructs. (e) The distance distribution function for VirA<sup>c</sup>DD/VirFG<sup>h</sup>DD (non-covalent complex) calculated with GNOM. The Guinier plot is inset.

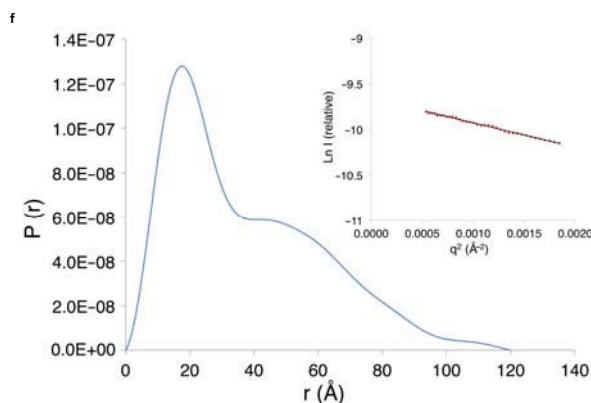


Figure S8 | Analysis by SAXS of the various constructs. (f) The distance distribution function for VirA ACP<sub>35</sub>-<sup>c</sup>DD/VirFG<sup>h</sup>DD (non-covalent complex) calculated with GNOM. The Guinier plot is inset.

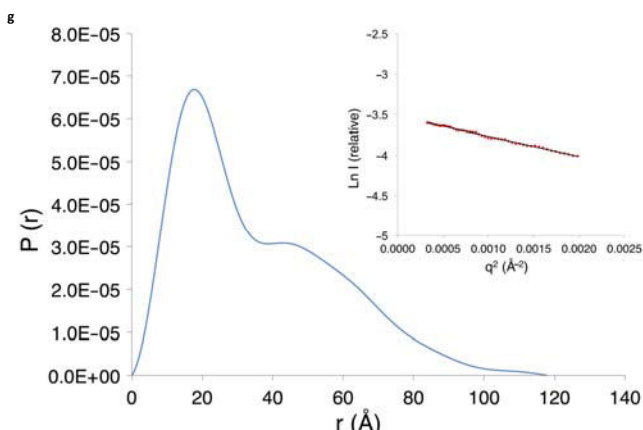


Figure S8 | Analysis by SAXS of the various constructs. (g) The distance distribution function for VirA ACP<sub>5b</sub>-<sup>5</sup>DD-VirFG<sup>5</sup>DD (covalent complex) calculated with GNOM. The Guinier plot is inset.

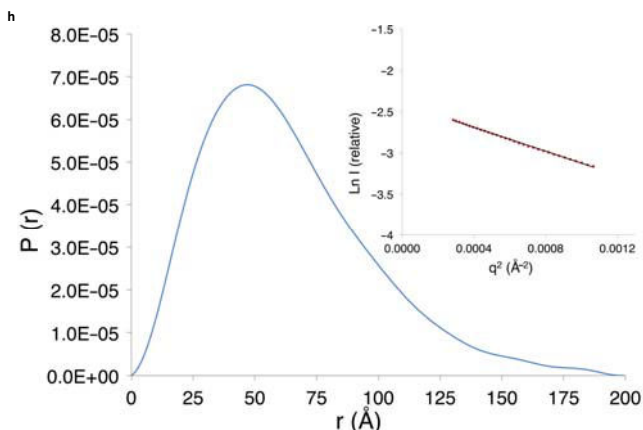


Figure S8 | Analysis by SAXS of the various constructs. (h) The distance distribution function for VirA ACP<sub>5b</sub>-<sup>5</sup>DD-VirFG<sup>5</sup>DD-KS<sub>6</sub> (covalent complex) calculated with GNOM. The Guinier plot is inset.

Table S5 | NMR and refinement statistics for VirA<sup>5</sup>DD-VirFG<sup>5</sup>DD.

Protein	
VirA <sup>5</sup> DD-VirFG <sup>5</sup> DD	
<b>NMR distance and dihedral constraints</b>	
Distance constraints	
Total NOE	1372
Intra-residue ( i-j  = 0)	338
Inter-residue ( i-j  ≠ 0)	1034
Sequential ( i-j  = 1)	371
Medium-range ( i-j  < 4)	333
Long-range ( i-j  > 5)	330
Hydrogen bonds	0
Total dihedral angle restraints	98
φ	49
ψ	49
<b>Structure statistics</b>	
Violations (mean and s.d.)	
Distance constraints (Å)	0.105 ± 0.008
Max. distance constraint violation (Å)	0.12
Dihedral angle constraints (°)	0 ± 0
Max. dihedral angle violation (°)	0
Deviations from idealized geometry	
Bond lengths (Å)	0.011 ± 0.001
Bond angles (°)	1.947 ± 0.024
Impropers (°)	n.d.
Average pairwise r.m.s. deviation* (Å)	
VirA <sup>5</sup> DD 6936-6960,	
VirFG <sup>5</sup> DD 3-25	
Heavy	0.99 ± 0.13
Backbone	0.38 ± 0.09

\*Pairwise r.m.s. deviation was calculated using 20 refined structures. n.d. = not determined.

Interface	C-terminal docking domain →										N-terminal docking domain →													
	6939	6940	6943	6945	6946	6949	6951	6953	6954	6956	6957	6960	3	4	6	7	10	11	15	17	19	20	23	24
VirA/FG	L	D	L	A	V	N	L	V	E	A	V	L	A	V	L	A	V	L	R	E	A	L	L	L
SnaE/E3	L	D	L	A	V	N	L	V	E	A	V	L	A	V	L	A	V	L	R	E	A	L	L	L
ChiE/F	L	A	V	R	M	-	-	-	A	V	E	L	D	Q	L	M	A	K	-	-	L	S	L	L
ChiC/D	L	W	L	A	L	-	-	A	E	D	D	L	R	Y	M	E	V	T	R	R	-	L	K	V
OzmN/H	I	E	V	S	L	-	-	T	V	D	D	L	A	R	L	L	-	-	-	-	-	-	-	L
OzmK/J	L	I	L	R	F	G	L	A	A	V	L	L	D	R	I	L	V	A	A	S	R	A	L	
OzmK/J	L	V	I	E	L	-	-	D	V	A	L	R	E	L	R	K	-	T	Q	A	Q	R	L	
TadJ/E	A	S	S	K	-	-	A	L	A	L	D	K	I	L	Y	R	T	T	Q	A	V	L	L	
TadJ/M	L	D	L	R	V	G	L	V	E	A	E	L	N	R	V	L	L	K	P	P	K	A	M	
5925/5924	L	D	L	E	I	L	-	D	D	L	A	L	S	L	R	L	L	S	L	L	A	L	R	
5923/5922	L	F	A	A	A	Q	A	A	A	L	A	I	D	K	L	L	S	I	I	K	L	R	L	
5922/5921	A	G	V	D	E	R	-	L	Q	M	Q	L	E	R	L	L	R	I	I	I	G	K	I	
5921/5920	P	A	L	Q	L	-	-	G	L	A	V	A	A	K	L	L	S	I	I	I	D	G	R	

Key (as observed in the VirA/VirFG complex):  
**Green/Orange** = inter-docking domain hydrophobic interface  
**Blue/Blue** = intra-docking domain hydrophobic interface  
**Pink/Pink/Purple** = inter-domain electrostatic interactions  
 Residue 6956 participates in two types of interactions, as indicated.

Supplementary Methods

Table S7   Primers used in this study.		
Constructs	Primer type	Oligonucleotides (5'-3')
VirA <sup>5</sup> DD	f	TTTCCGGGATCTATACCGGCCCGGGAGCCCTC
	r	TTTCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
VirFG <sup>5</sup> DD	f	TTTCCGGATCCATGGACGCGAAAGAGATCCTGACCCGATCAAG
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
VirFG <sup>5</sup> DD-linker	f	TTTCCGGATCCATGGACGCGAAAGAGATCCTGACCCGATCAAG
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
VirA ACP <sub>5b</sub> - <sup>5</sup> DD	f	TTTCCGGATCCGCGTGGCCGTGACCCCG
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
VirFG <sup>5</sup> DD-KS <sub>6</sub>	f	TTTCCGGATCCATGGACGCGAAAGAGATCCTGACCCGATCAAG
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
BamHI f		GACACGTTTTCGCGAGCGCATCTCCGACGGCGAC
	BamHI r	GTCCGCGTCCGAGGATGCGCTGCCAGAACGCTGTC
VirA <sup>5</sup> DD-VirFG <sup>5</sup> DD (covalent fusion)	f	TTTCCGGATCTATACCGGCCCGGGAGCCCTC
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
Overlap1		GGAGGTAGTGGTGGAGGCGAGTATGGACGCGAAAGAGATCCTGACCCGATCAAG
	Overlap2	GCCTCCACCACTACCTCCCGCGCCGGGGGTGAGCAG
VirA ACP <sub>5b</sub> - <sup>5</sup> DD-VirFG <sup>5</sup> DD (covalent fusion)	f	TTTCCGGATCCGCGTGGCCGTGACCCCG
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
Overlap1		GGAGGTAGTGGTGGAGGCGAGTATGGACGCGAAAGAGATCCTGACCCGATCAAG
	Overlap2	GCCTCCACCACTACCTCCCGCGCCGGGGGTGAGCAG
VirA ACP <sub>5b</sub> - <sup>5</sup> DD-VirFG <sup>5</sup> DD-KS <sub>6</sub> (covalent fusion)	f	TTTCCGGATCCGCGTGGCCGTGACCCCG
	r	TTTCCGAAAGCTTTCAGCGCCCGGGGTGAGCAG
VirA <sup>5</sup> DD D6940A	f	TCCAGGCGACCTGGCCGCGTCTCAGCGCC
	r	GGCGTGGACGCGCGCCAGGTCCGCTGGGA
VirA <sup>5</sup> DD D6940R	f	TCCAGGCGACCTGGCCGCGTCTCAGCGCC
	r	GGCGTGGACGCGCGCCAGGTCCGCTGGGA
VirA <sup>5</sup> DD L6943A	f	GACCTGGACGCTGCCAGGCGCGTCCGCGAC
	r	GTCCGCGACGCGCGTCCGCGCGTCCGCGAC
VirA <sup>5</sup> DD N6949A	f	AGCCGCGTCCGCGACGCGCGCTGTCCATCGAG
	r	CTCGATGGACGCGCGCGTCCGCGCGCGCT
VirA <sup>5</sup> DD N6949K	f	AGCCGCGTCCGCGACAAACGCGTGTCCATCGAG
	r	CTCGATGGACGCGCTTGTCCGCGCGCGCT
VirA <sup>5</sup> DD E6954A	f	AACCGGCTGTCCATCGCGAGCGCGTCAACCTG
	r	CAGGGTGAACCGCTCCGCGATGGACAGCGGTT
VirA <sup>5</sup> DD E6954K	f	AACCGGCTGTCCATCAAGCAGCGCGTCAACCTG
	r	CAGGGTGAACCGCTCCGCGATGGACAGCGGTT
VirFG <sup>5</sup> DD K4A	f	GGATCCATGGACGCGGAGAGATCCTGACCCGA
	r	TCCGCGTCCGCGTCCGCGTCCATGGATCC
VirFG <sup>5</sup> DD K4E	f	GGATCCATGGACGCGGAGAGATCCTGACCCGA
	r	TCCGCGTCCGCGTCCGCGTCCATGGATCC
VirFG <sup>5</sup> DD K11A	f	ATCCTGACCCGATTCCGCGAGCGGGGCGTGGAC
	r	GTCCAGGCCCGCTCCGCGAATCGGGTCAGGAT
VirFG <sup>5</sup> DD R17A	f	GACGGGGCGCTGACGCGCGCCCGCCAGGCG
	r	CGCTGGCGCGCGCGCTCCAGGCCCGCTG
VirFG <sup>5</sup> DD R17D	f	GACGGGGCGCTGACGCGCGCCCGCCAGGCG
	r	CGCTGGCGCGCGCTCCAGGCCCGCTG

<sup>a</sup>Introduced restriction sites are underlined, while the added Tyr is indicated by bold letters.

Supplementary References

- Whicher, J.R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
- Davidson, J. *et al.* Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem. Sci.* **5**, 3081–3095 (2014).
- Chang, Z. *et al.* Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **67**, 1356–1367 (2004).
- Partida-Martínez, L.P. & Hertweck, C. A gene cluster encoding rhizoxin biosynthesis in *Burkholderia rhizoxina*, the bacterial endosymbiont of the fungus *Rhizopus microsporus*. *ChemBioChem* **8**, 41–45 (2007).
- El-Sayed, A.K. *et al.* Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419–430 (2003).
- Edwards, D.J. *et al.* Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004).
- Gaitatzis, N. *et al.* The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* **277**, 13082–13090 (2002).
- Nguyen, T. *et al.* Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
- Simunovic, V. *et al.* Myxovirescin A biosynthesis is directed by hybrid polyketide synthase/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. *ChemBioChem* **7**, 1206–1220 (2006).
- Buntin, K. *et al.* Production of the antifungal isochromanone ajundazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed Engl.* **47**, 4595–4599 (2008).
- Dehn, R. *et al.* Molecular basis of elansolid biosynthesis: evidence for an unprecedented quinone methide initiated intramolecular Diels-Alder cycloaddition/macrolactonization. *Angew. Chem. Int. Ed Engl.* **50**, 3882–3887 (2011).
- Pulsawat, N., Kitani, S. & Nihira, T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **393**, 31–42 (2007).
- Mast, Y. *et al.* Characterization of the ‘pristinamycin supercluster’ of *Streptomyces pristinaespiralis*. *Microb. Biotechnol.* **4**, 192–206 (2011).
- Perlova, O., Gerth, K., Kaiser, O., Hans, A. & Müller, R. Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. *J. Biotechnol.* **121**, 174–191 (2006).
- Erol, O. *et al.* Biosynthesis of the myxobacterial antibiotic coralopyronin A. *ChemBioChem* **11**, 1253–1265 (2010).
- Zhang, F. *et al.* Cloning and elucidation of the FR901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate as starter units. *J. Am. Chem. Soc.* **133**, 2452–2462 (2011).
- Weber, T. *et al.* Molecular analysis of the kirromycin biosynthetic gene cluster revealed  $\beta$ -alanine as precursor of the pyridone moiety. *Chem. Biol.* **15**, 175–188 (2008).
- Matilla, M.A., Stöckmann, H., Leeper, F.J. & Salmond, G.P. C. Bacterial biosynthetic gene clusters encoding the anti-cancer heteromalidic class of molecules: biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *J. Biol. Chem.* **287**, 39125–39138 (2012).
- Zhao, C. *et al.* Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an ‘acyltransferase-less’ type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem.* **285**, 20097–20108 (2010).
- Wang, C. *et al.* Thailandepsins: bacterial products with potent histone deacetylase inhibitory activities and broad-spectrum antiproliferative activities. *J. Nat. Prod.* **74**, 2031–2038 (2011).
- Fukuda, D. *et al.* A natural plasmid uniquely encodes two biosynthetic pathways creating a potent anti-MRSA antibiotic. *PLoS One* **6**, e18031 (2011).
- Mahenthalingam, E. *et al.* Encyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* genomic island. *Chem. Biol.* **18**, 665–677 (2011).
- Wakimoto, T. *et al.* Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat. Chem. Biol.* **10**, 648–655 (2014).
- Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
- Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
- Svergun, D., Barberato, C. & Koch, M. CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
- Dosztányi, Z., Csizmek, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
- Dosztányi, Z., Mészáros, B. & Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746 (2009).
- Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).
- Li, X., Romero, P., Rani, M., Dunker, A.K. & Obradovic, Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform. Workshop Genome Inform.* **10**, 30–40 (1999).
- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645 (2004).
- Linding, R. *et al.* Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453–1459 (2003).
- Pavlović-Lažetić, G.M. *et al.* Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics* **12**, 66 (2011).
- Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
- Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F. & Weissman, K.J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
- Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
- Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
- Kampa, A. *et al.* Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3129–E3137 (2013).
- Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16222–16227 (2004).
- Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14002–14007 (2002).
- Fisch, K.M. *et al.* Polyketide assembly lines of uncultivated sponge symbioses from structure-based gene targeting. *Nat. Chem. Biol.* **5**, 494–501 (2009).
- Sudek, S. *et al.* Identification of the putative bromystatin polyketide synthase gene cluster from *Candidatus Endobugula sertula*, the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J. Nat. Prod.* **70**, 67–74 (2007).
- Menche, D. *et al.* Stereochemical determination and complex biosynthetic assembly of etnangien, a highly potent RNA polymerase inhibitor from the myxobacterium *Sorangium cellulosum*. *J. Am. Chem. Soc.* **130**, 14234–14243 (2008).
- Butcher, R.A. *et al.* The identification of bacillaene, the product of the PksX megacassette in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1506–1509 (2007).
- Mattheus, W. *et al.* Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. *Chem. Biol.* **17**, 149–159 (2010).
- Carvalho, R., Reid, R., Viswanathan, N., Gramajo, H. & Julien, B. The biosynthetic genes for disorazole, potent cytotoxic compounds that disrupt microtubule formation. *Gene* **359**, 91–98 (2005).
- Kopp, M., Irschik, H., Pradella, S. & Müller, R. Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. *ChemBioChem* **6**, 1277–1286 (2005).
- Pistorius, D. & Müller, R. Discovery of the rhizopodin biosynthetic gene cluster in *Stigmatella aurantiaca* Sg a15 by genome mining. *ChemBioChem* **13**, 416–426 (2012).
- Chen, X.-H. *et al.* Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* **188**, 4024–4036 (2006).
- Cheng, Y.-Q., Tang, G.-L. & Shen, B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–3154 (2003).
- Schneider, K. *et al.* Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J. Nat. Prod.* **70**, 1417–1423 (2007).
- Irschik, H. *et al.* Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrosynthetic analysis for deciphering natural product assembly by *trans-AT* PKS. *ChemBioChem* **11**, 1840–1849 (2010).
- Liu, X. *et al.* Genomics-guided discovery of thailanstatins A, B, and C as pre-mRNA splicing inhibitors and antiproliferative agents from *Burkholderia thailandensis* MSMB43. *J. Nat. Prod.* **76**, 685–693 (2013).
- Moebius, N. *et al.* Biosynthesis of the respiratory toxin bongkrekic acid in the pathogenic bacterium *Burkholderia gladioli*. *Chem. Biol.* **19**, 1164–1174 (2012).
- Seo, J.-K. *et al.* Comparative characterization of the lactimidomycin and iso-migrastatin biosynthetic machineries revealing unusual features for acyltransferase-less type I polyketide synthases and providing an opportunity to engineer new analogues. *Biochemistry* **53**, 7854–7865 (2014).
- Arakawa, K., Sugino, F., Kodama, K., Ishii, T. & Kinashi, H. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* **12**, 249–256 (2005).
- Lim, S.-K. *et al.* Migrastatin, migrastatin, and dorrigin production in *Streptomyces platensis* NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *J. Biol. Chem.* **284**, 29746–29756 (2009).
- Royer, M. *et al.* Albicidin pathotoxin produced by *Xanthomonas albilineans* is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes. *Mol. Plant-Microbe Interact.* **17**, 414–427 (2004).
- Ross, A.C. *et al.* Biosynthetic multisteps facilitates thalassospiramide structural diversity in marine bacteria. *J. Am. Chem. Soc.* **135**, 1155–1162 (2013).
- McGuffin, L.J., Bryson, K. & Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

Supplementary Note

We aimed to strengthen the case that our observations concerning the VirA<sup>CD</sup>/VirFG<sup>DD</sup> interface are generalizable to other *trans-AT* PKSs. For this, we benchmarked several web-based servers for the *in silico* identification of IDPs (IUPred/ANCHOR<sup>51,52</sup>, PONDR<sup>53,54</sup>, DisoPred<sup>55</sup> and DisEMBL<sup>56</sup>) against VirA<sup>CD</sup> and VirFG<sup>DD</sup>, to determine if they accurately predicted their respective behaviors. These approaches have complementary principles of operation (they are based either on alignments of homologous protein sequences or the physicochemical properties of amino acids in proteins<sup>57</sup>) and so a combination of programs should give a more confident prediction. We also analyzed the sequences using a coiled-coil predictor<sup>58</sup>, which we validated against the known coiled-coil regions of several Class 1 N-terminal DDs<sup>59,60</sup>. Among these programs, IUPred/ANCHOR and PONDR yielded prediction data in greatest agreement with the observed behavior of the Vir docking domains. Both programs correctly predicted that the overall VirA<sup>CD</sup> interaction region (residues Q6936–L6960) is largely structured, while ANCHOR identified the two  $\alpha$ -helices as binding motifs (data provided on pages 42–43). Furthermore, no coiled-coil was predicted for the sequence. In the case of VirFG<sup>DD</sup> (residues A3–A25), both IUPred/ANCHOR and PONDR predicted a sequence on the borderline between structured and disordered, while ANCHOR identified the whole sequence as a binding region. Again, the coiled-coil prediction was negative. Gratifyingly, analogous results were obtained for the regions corresponding to VirA<sup>CD</sup> and VirFG<sup>DD</sup> from the homologous pristinamycin PKS (SnaE2<sup>CD</sup> and SnaE3<sup>DD</sup>, respectively), despite overall low levels of sequence identity between the docking regions (53% between VirA<sup>CD</sup> and SnaE2<sup>CD</sup> and 39% between VirFG<sup>DD</sup> and SnaE3<sup>DD</sup>) (pages 44–45).

We next extended this analysis to the remaining 14 putative VirA<sup>CD</sup>/VirFG<sup>DD</sup>-type docking interfaces from *trans-AT* PKSs assigned on the basis of sequence analysis and secondary structure prediction (data provided on pages 46–79). For nine of these, the predictions were in strong agreement with those for VirA<sup>CD</sup> and VirFG<sup>DD</sup>, although analysis by IUPred indicated a relatively higher degree of disorder in certain C-terminal DDs than for VirA<sup>CD</sup> itself, while TaiE appears significantly more ordered than VirFG<sup>DD</sup>. For the five interfaces which clearly did not conform (KirAV/KirAVI, 5923/5922, 5922/5921, 5921/5920 and RizC/RizD), the prediction was instead for a coiled-coil region within the N-terminal DDs. It is perhaps notable that three of these interfaces occur within the encyloxin system, where as a result of the same class of docking domains operating at four junctions (all of which are ACP/KS interfaces) the specificity demands are presumably higher.

Couples analyzed:

VirA/VirFG  
 SnaE2/SnaE3  
 KirAlI/KirAlII  
 KirAV/KirAVI  
 TaiL/TaiM  
 TaiD/TaiE  
 5925/5924  
 5922/5921  
 5921/5920  
 5923/5922  
 OzmK/L  
 OzmH/J  
 OzmN/H 1  
 ChiC/ChiD  
 ChiE/ChiF  
 DisB/DisC  
 RizC/RizD

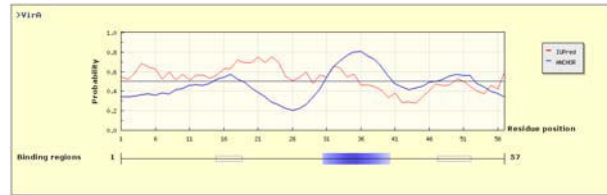
Key:

Red = For VirA<sup>CD</sup>/VirFG<sup>DD</sup> observed  $\alpha$ -helices in the complex; for all others,  $\alpha$ -helices predicted by PsiPred  
 Blue = start/end site of expression constructs for VirA<sup>CD</sup>/VirFG<sup>DD</sup>, respectively  
 All sequences are numbered starting from 1 to match the output of the programs

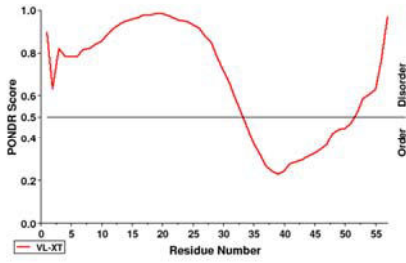
VirA<sup>Δ</sup>DD

AGRAAPVFAAAAEPALPLPVITGAGEPSQADLDALLSAVRDNRRLSIEQAVTLTPRR

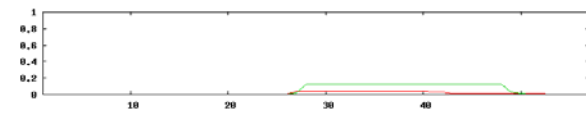
IUPred/ANCHOR



PONDR



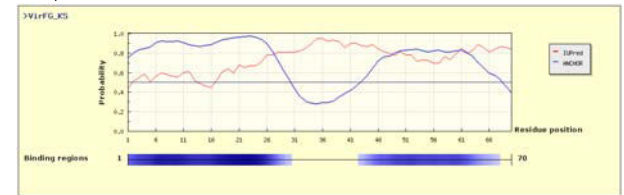
COILED-COIL PREDICTOR



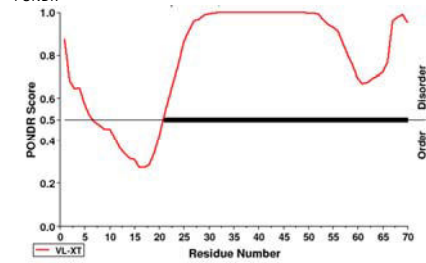
VirFG<sup>Δ</sup>DD

MDAKEILTRFKDGGDLDRAAAQALLAGRTPAAAPRSEPAAPTRPAVPAVPAVEPAAGTTVAEETAGRP

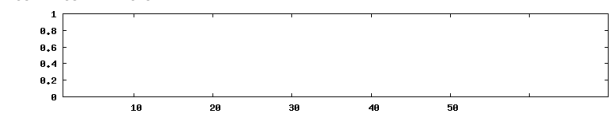
IUPred/ANCHOR



PONDR



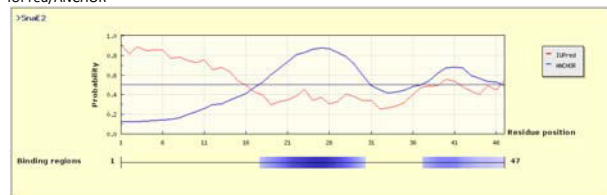
COILED-COIL PREDICTOR



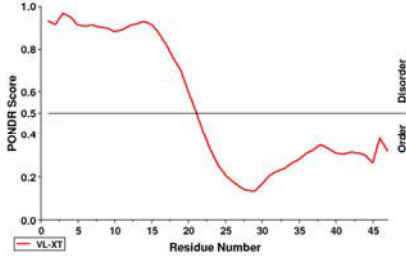
SnaE2<sup>Δ</sup>DD

TSRTAPQFAGAVPAGSVSAADLDALLAAVRDNRRLTVEQALALLPQHT

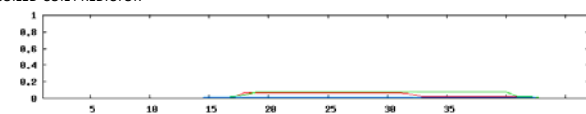
IUPred/ANCHOR



PONDR



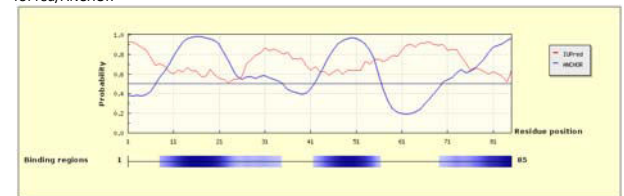
COILED-COIL PREDICTOR



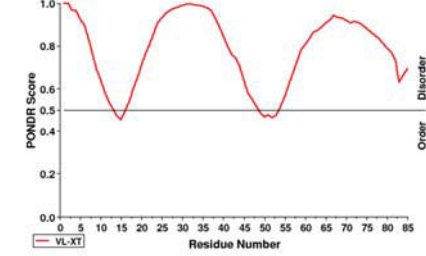
SnaE3<sup>Δ</sup>DD

MPAEPD<sup>Δ</sup>VDRDRAAALLRAVAGQ<sup>Δ</sup>RGFPAAGAATASARPA<sup>Δ</sup>PGGVDR<sup>Δ</sup>EA<sup>Δ</sup>AALLRAA<sup>Δ</sup>AGERRP<sup>Δ</sup>GGDGPAGGSMGRTLP<sup>Δ</sup>EA<sup>Δ</sup>GSAGRVL

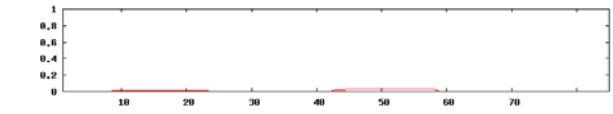
IUPred/ANCHOR



PONDR



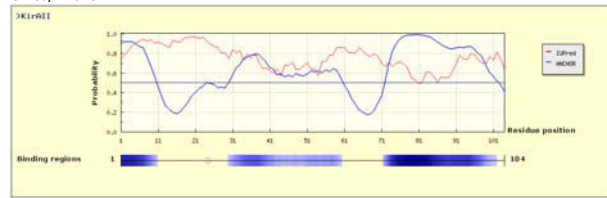
COILED-COIL PREDICTOR



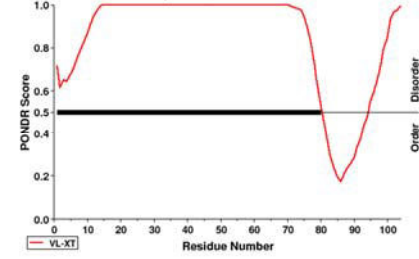
KirAII<sup>ΔDD</sup>

HDKYGLPATATPEPPAPVTSSATAPEPTSVFVASSAAAPAPIPAPVAPSAAPAPVASSAAPEPPAPVAADA  
 VGALLAGLRDGSVTVDDALAE LRQGNVR  
 72 85 91 99

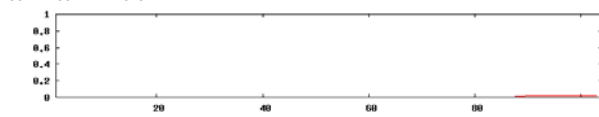
IUPred/ANCHOR



PONDR



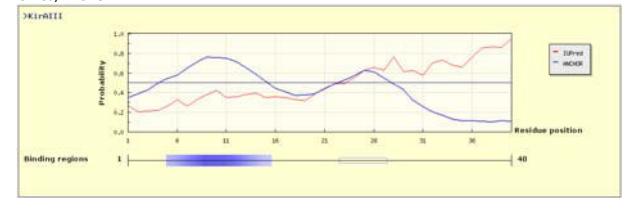
COILED-COIL PREDICTOR



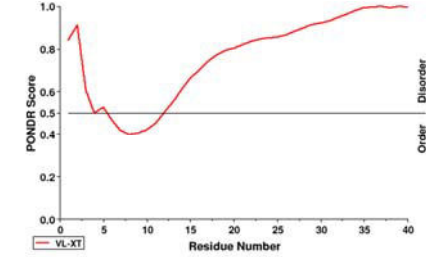
KirAIII<sup>ΔDD</sup>

3 13 18 29  
 MSRALRDVLAAVAAGELPDEIAESLLRGLSDPVPEPGYP

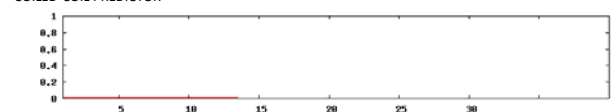
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR



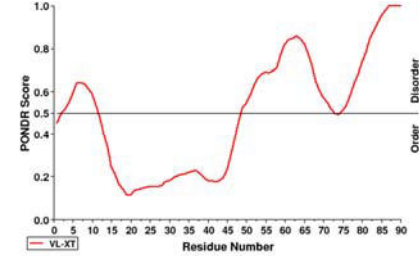
KirAV<sup>ΔDD</sup>

AGRFDGGTGGADLLPALFADFRLAAGLTEAA TDEVFRDRVVERVRLLTALSPATPEAPPDLFAAATDDEVFALI  
 DAELGSPVRGQEDR  
 71 79

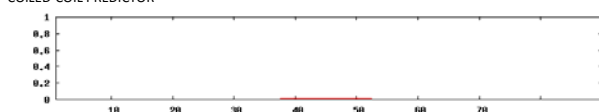
IUPred/ANCHOR



PONDR



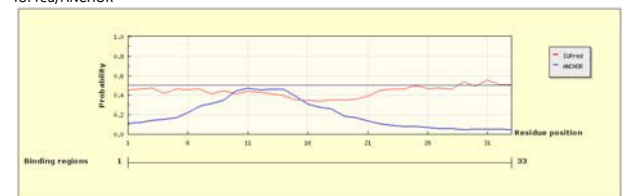
COILED-COIL PREDICTOR



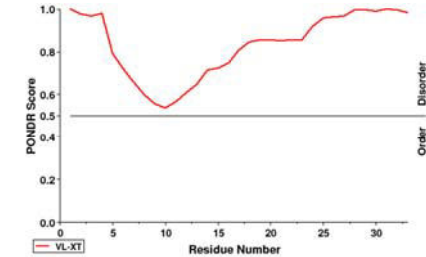
KirAVI<sup>ΔDD</sup>

4 31  
 MSSDKLRDYLRRVTAE LQVTRRRLEVESRSG

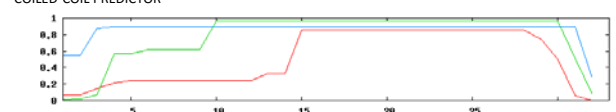
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR







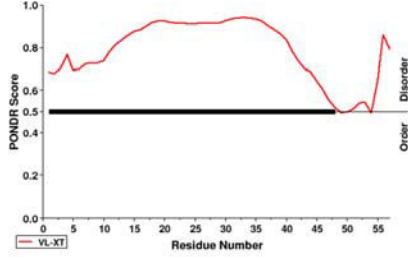
5925 <sup>DD</sup>

LAANAPAGSAVAEEAASVAPAVEATEPEAAAPLDTELSEIEGLQDDDLAALLGKEFIRE

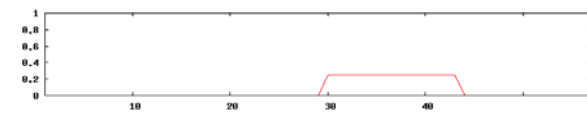
IUPred/ANCHOR



PONDR



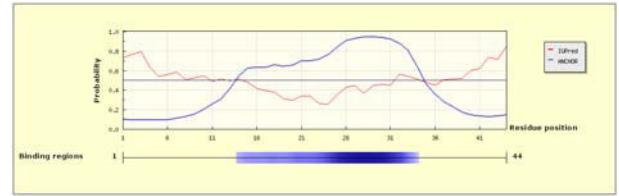
COILED-COIL PREDICTOR



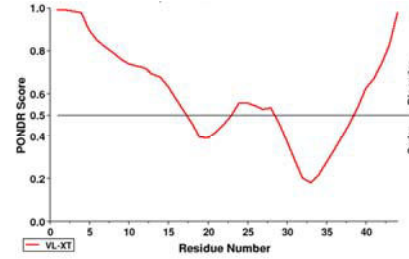
5924 <sup>DD</sup>

MNKPTSSDGWKDDYLSRLSRLSKNQLMALALKKQQQLPQPAA

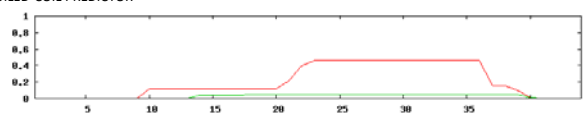
IUPred/ANCHOR



PONDR



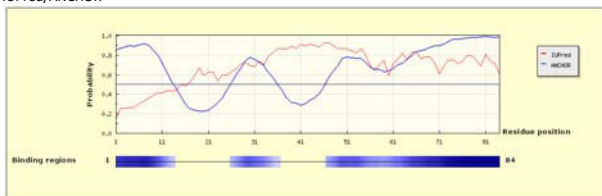
COILED-COIL PREDICTOR



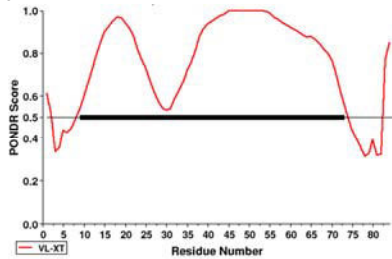
5922 <sup>DD</sup>

AARLKALYAAAPAAAGSTGNAGATGATSAFSAADATHPDATDAPPFAGPLAATVSPPTASDAAGDAVTDEIERELQTMQALLIEDR

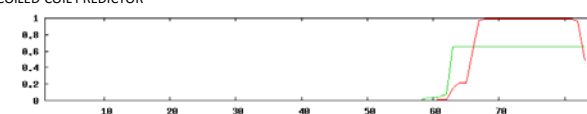
IUPred/ANCHOR



PONDR



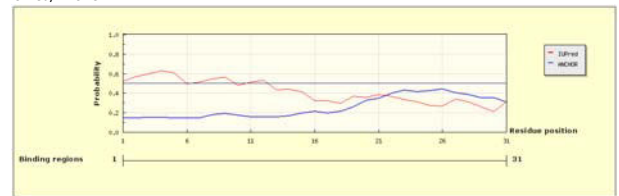
COILED-COIL PREDICTOR



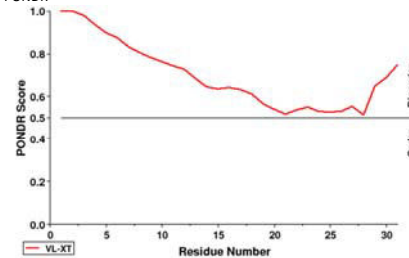
5921 <sup>DD</sup>

MDSKDRELLARQRDAIRQGIQKIQALSARLD

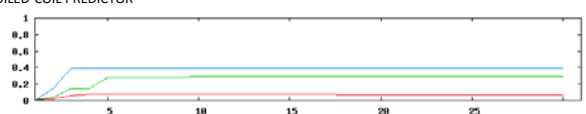
IUPred/ANCHOR



PONDR



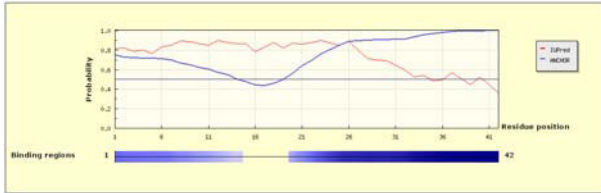
COILED-COIL PREDICTOR



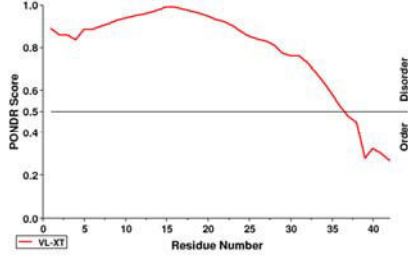
5921 <sup>o</sup>DD

AGPAQAARAPASASAPTFRRAAPADALDQLDEGELADVLDKLL

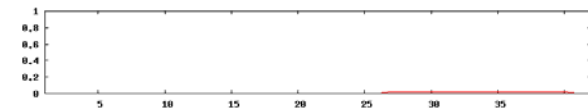
IUPred/ANCHOR



PONDR



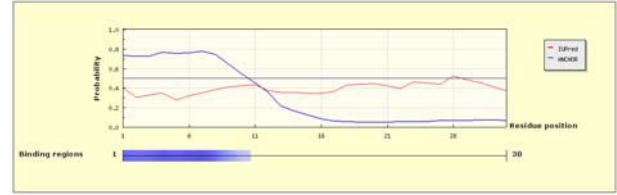
COILED-COIL PREDICTOR



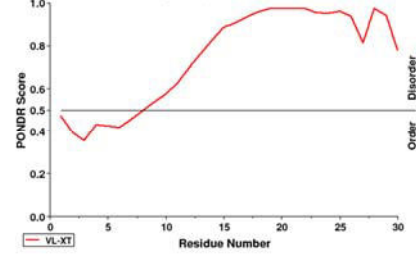
5920 <sup>o</sup>DD

MNDYKALLKSSIRKIQQDRRIRELESGVD

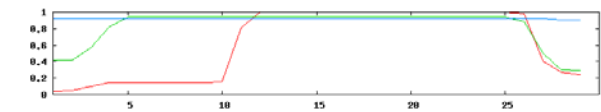
IUPred/ANCHOR



PONDR



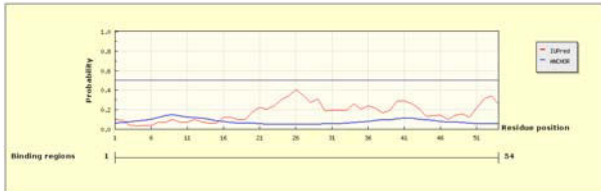
COILED-COIL PREDICTOR



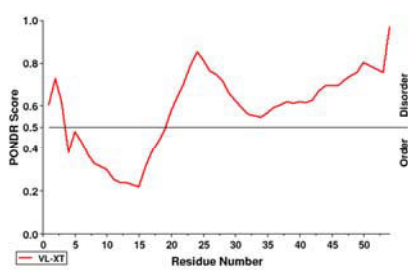
5923 <sup>o</sup>DD

ARGALAEIIFAPAIVAAPQLANAADASIIGDHSAAELARILAEHLGGLESRGAL

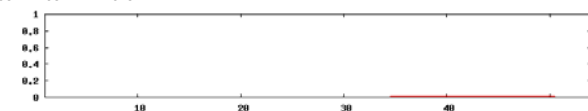
IUPred/ANCHOR



PONDR



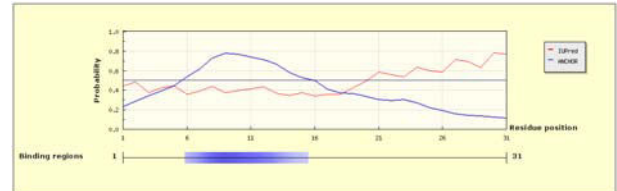
COILED-COIL PREDICTOR



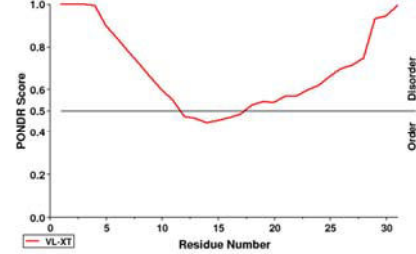
5922 <sup>o</sup>DD

MTMDKDLLLQSIQTIRELKTRLAQAEQGHH

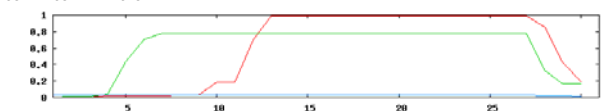
IUPred/ANCHOR



PONDR

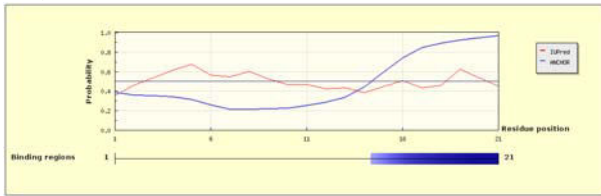


COILED-COIL PREDICTOR



OzmK<sup>h</sup>DD  
 2 8 12 19  
 EAE LVERIEELPEDTVATLLA

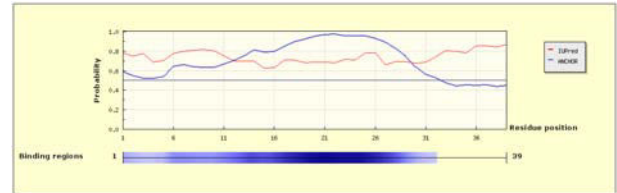
IUPred/ANCHOR



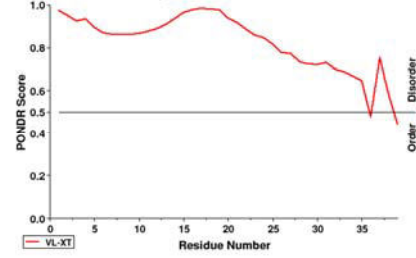
Sequence is too short for PONDR/Coiled-coil predictor

OzmL<sup>h</sup>DD  
 3 13 16 29  
 MSREDLATSRKARLSPQKRALLEKLSGAARTSGTTTRIPR

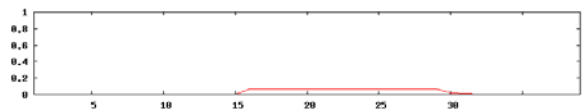
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR

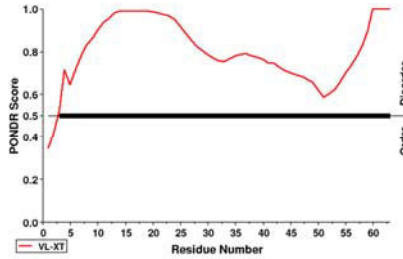


OzmH<sup>h</sup>DD  
 28 43 46 60  
 IELVGFDEPTAGSGEPPADGTGDTAAPEDELI AKLEARFAAGELSAAEVL D LLD AELATRRER

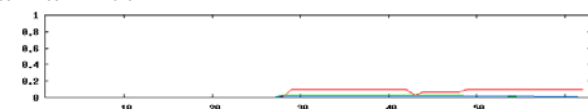
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR

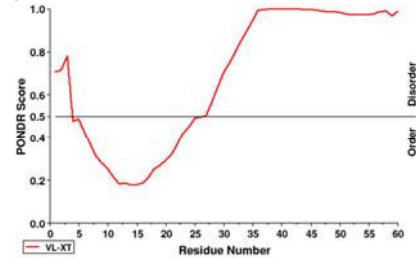


OzmJ<sup>h</sup>DD  
 3 16 20 29  
 MSTAEFRDILDYARGAVGAEQARDLLAARRPAQAPGRPTAPAPSGQDAATGPADRPETL

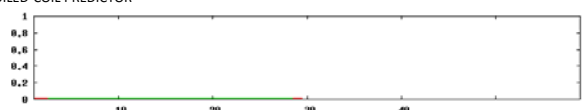
IUPred/ANCHOR



PONDR



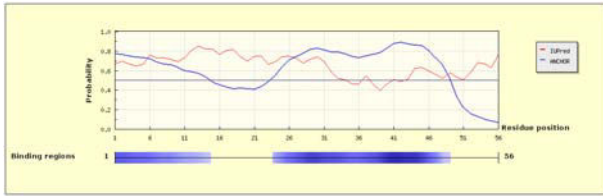
COILED-COIL PREDICTOR



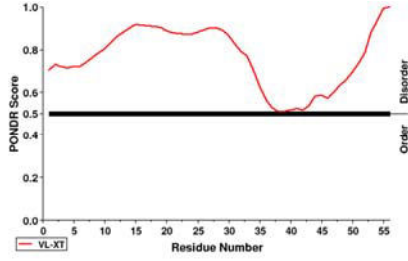
OzmN<sup>c</sup>DD

22 54  
 AATGAAGTAAGTAAGPARAAEPEAGIEQLVDSLSPTEVDSLRLRQLGSLVLRQKEEQR

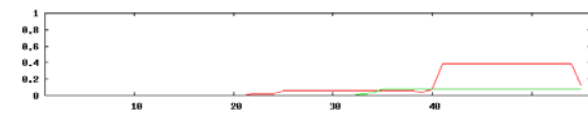
IUPred/ANCHOR



PONDR



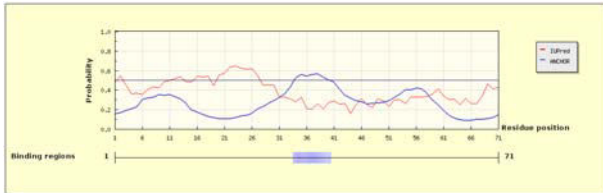
COILED-COIL PREDICTOR



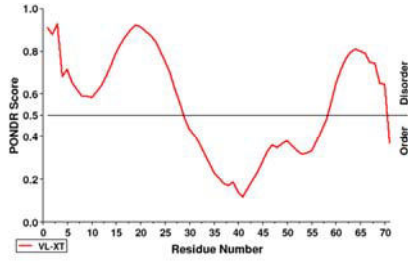
ChiC<sup>c</sup>DD

31 62 65 68  
 ERDPALAEIKAQVLRAGAPGPAPELAVGHDELWRLLGALGDAELDALGDAELDALFNALRSRPELEGMV

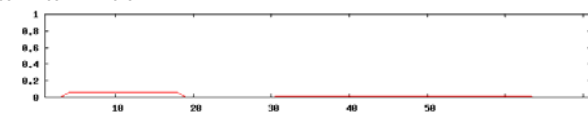
IUPred/ANCHOR



PONDR



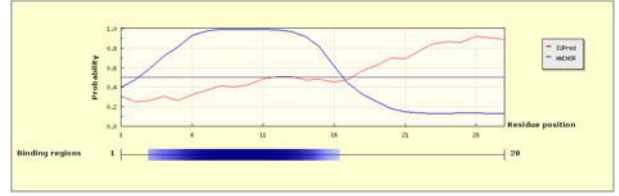
COILED-COIL PREDICTOR



OzmH<sup>n</sup>DD

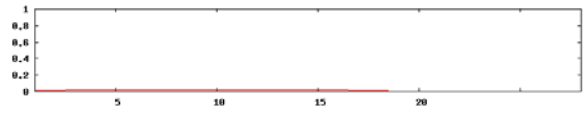
6 24  
 MRRRSMDKRALLLKLLASVRQQAAPATAPRRTA

IUPred/ANCHOR



Too short for PONDR

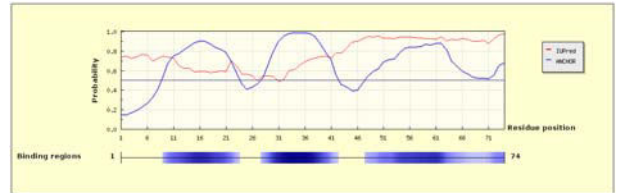
COILED-COIL PREDICTOR



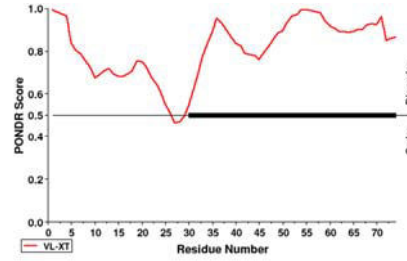
ChiD<sup>n</sup>DD

5 43  
 MQRDPDRGRRLLYRMEKAVTDMFDRSREEKLSVFRALRDRLRRTPPQHGAADGAGPRAPVLAAPDPQNRHRPFP

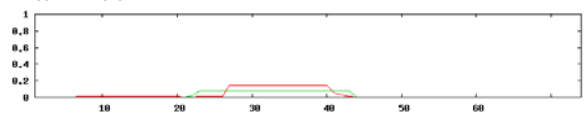
IUPred/ANCHOR



PONDR



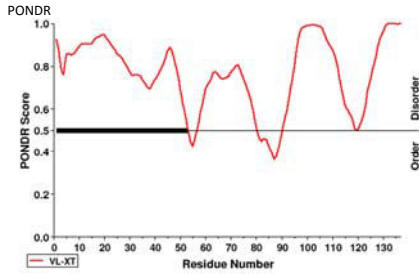
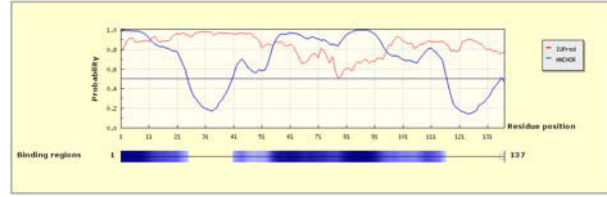
COILED-COIL PREDICTOR



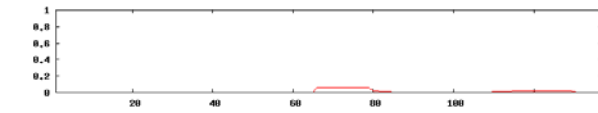
ChiE<sup>Δ</sup>DD

VDHHEAALRASEPPVGGGGAEFVRPRANPERPERADPPERAGVAPGRRAPSPFAAAVERFDGAVPEDLAGEVARM  
 SDAEVE<sup>RL</sup>LLGWLQGGGAGGAGGAGATNGSSAVRASLDDGQREAE<sup>RR</sup>RLDRGSPGSEWR  
 79 99 105 108 119

IUPred/ANCHOR



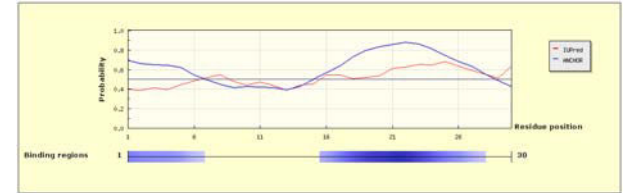
COILED-COIL PREDICTOR



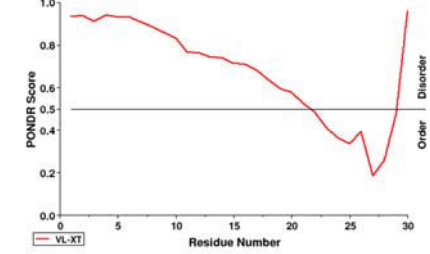
ChiF<sup>Δ</sup>DD

MAMTLRSQDL<sup>MD</sup>DAKRESL<sup>VR</sup>LLHQK<sup>KA</sup>NR  
 5 28

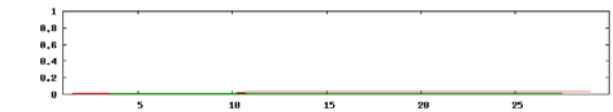
IUPred/ANCHOR



PONDR



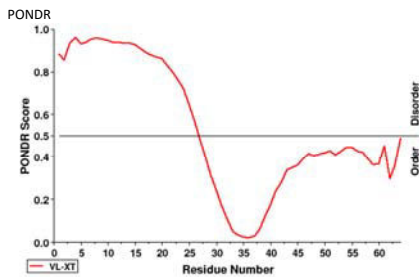
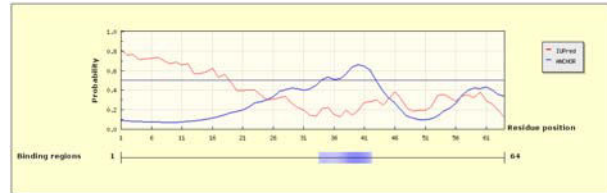
COILED-COIL PREDICTOR



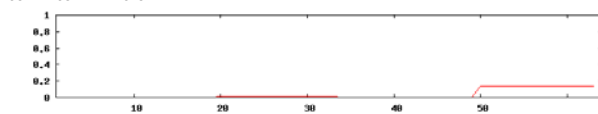
DisB<sup>Δ</sup>DD

SRRGARAPESRAKARSGAA<sup>GADL</sup>STSLKALSGAVL<sup>REQFLAF</sup>GHDLAGVPG<sup>EEL</sup>TRLYAIL<sup>QEE</sup>  
 20 46 51 62

IUPred/ANCHOR



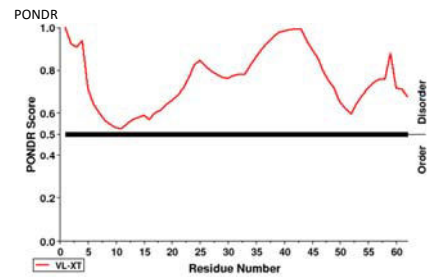
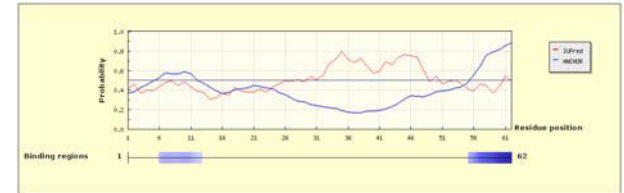
COILED-COIL PREDICTOR



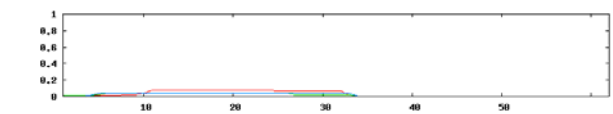
DisC<sup>Δ</sup>DD

MESAM<sup>TIQEF</sup>ANLSA<sup>E</sup>EKVQVLR<sup>LRDR</sup>RRASWQA<sup>A</sup>EPGPAASAQ<sup>P</sup>SLRPVITARP<sup>GDR</sup>FLP<sup>FP</sup>  
 6 11 15 30

IUPred/ANCHOR



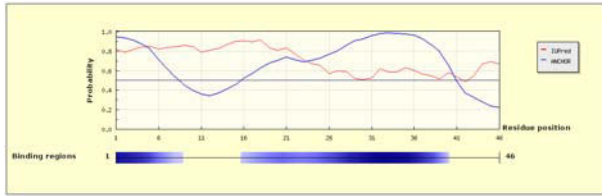
COILED-COIL PREDICTOR



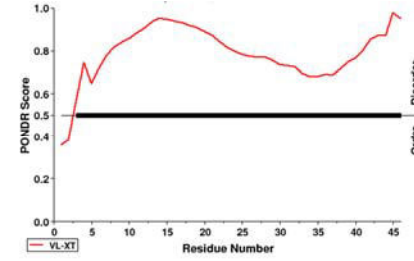
RizC<sup>DD</sup>

RALVTEAPPRGATQ SANGH A L E G M S D G N L E E L F A R L S G V S S Q K G E A

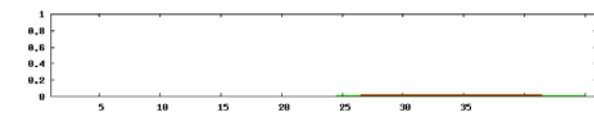
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR



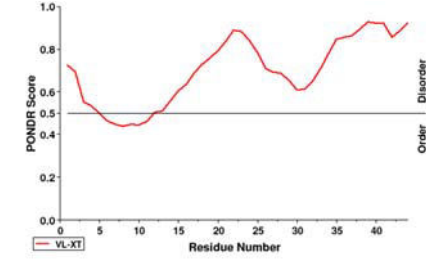
RizD<sup>DD</sup>

M P F K N D S A L N E F R L L S R E E K L Q R I A A L I E E K K S S A K S A S S E A F A

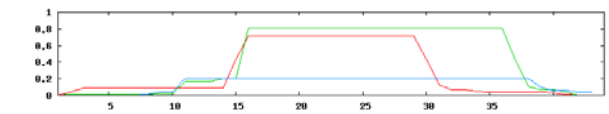
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR



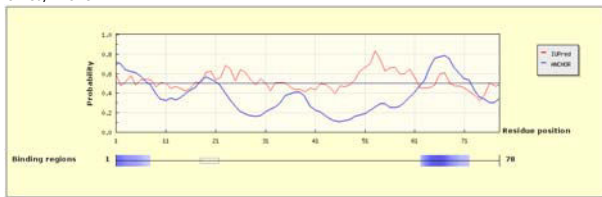
CONTROL: CLASS 1A DDs

Key:  
**Red** =  $\alpha$ -helices observed in the NMR structure of the DEBS 2<sup>DD</sup>/DEBS 3<sup>DD</sup> complex  
**Blue** =  $\alpha$ -helices predicted by PsiPred

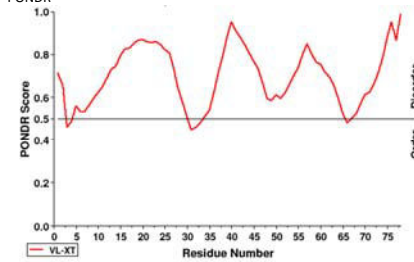
DEBS 2<sup>DD</sup>

A A S P A V D I G D R I D E L E R A L E A L S A E D G H D D V G Q R L E S L L R R W N S R R A D A P S T S A I S E D A S D D E L F S M L D R F G G G E

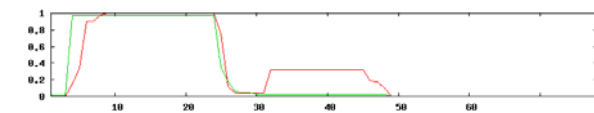
IUPred/ANCHOR



PONDR



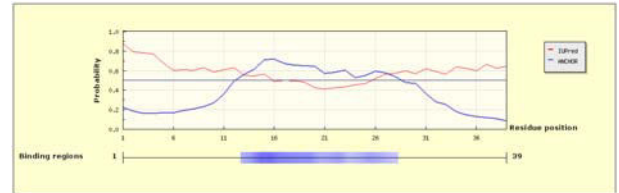
COILED-COIL PREDICTOR



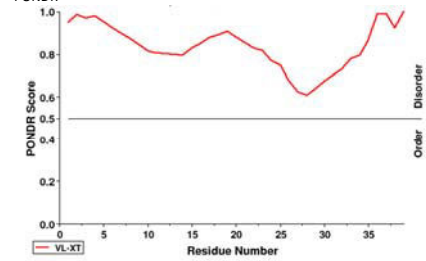
DEBS 3<sup>DD</sup>

M S G D N G M T E E K L R R Y L K R T V T E L D S V T A R L R E V E H R A G E

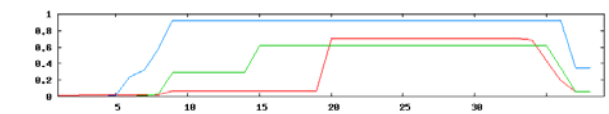
IUPred/ANCHOR



PONDR



COILED-COIL PREDICTOR



## Partie 2 : Conversion d'un module non-épimérisant en un module épimérisant par échange de domaine KR

(Annaval *et al.*, 2015)<sup>99</sup>

Notre objectif a été de transformer un module non épimérisant en épimérisant par échange de KR. L'ensemble des KR épimérisantes, provenant de modules ne contenant que des KR et pas d'autres domaines réducteurs (DH et ER), a remplacé la KR du module 2 (de type A1) de la PKS DEBS1-TE. La production de polykétide découlant des PKS résultantes s'est réalisée *in vivo*. Après extraction des polykétides, la stéréochimie a été analysée par GC-MS.

Nous avons identifié 16 KR dont les séquences ont été clonées dans le vecteur pMU3 contenant la séquence originale DEBS 1-TE provenant du plamide pJLK117. Les séquences des KR ont été insérées en premier lieu *via* deux sites de restriction différents en 5', *AvrII* et *BglII*, qui ont donné les meilleurs rendements en lactone lors de l'étude de Kellenberger *et al.*<sup>68</sup> et *via* le site *NheI* en 3'. 32 PKS différentes ont ainsi été créées. Cependant, au cours de ce travail, les résultats de Zheng *et al.*,<sup>100</sup> ont mis en évidence l'existence d'un domaine de dimérisation en amont de chaque domaine KR à l'exception des domaines KR des modules 1. Ainsi les séquences de ces 12 domaines KR ont été reclés afin d'inclure le domaine de dimérisation en Nter, mais cette fois-ci en n'utilisant qu'un seul site restriction en 5', *BglII*.

Les lactones ont été produites dans des cellules de *S. erythraea* JC2 qui est la souche délétée dans son génome de la région *eryA* qui code DEBS (à l'exception de la région de *eryAIII* qui code le domaine TE)<sup>36</sup>. Les lactones ont été extraites pour être ensuite analysées par GC-MS (collaboration C. Paris, LiBio, Université de Lorraine). Une méthode de GC a été développée afin que chaque molécule susceptible d'être produite (c'est-à-dire les lactones où la stéréochimie change au niveau des deux positions contrôlées par la KR (C2 et C3) ainsi que la lactone non-réduite en position C3) soit bien séparée, et permettre ainsi de mesurer son temps de rétention et le rendement de production.

Mise à part pour une construction, chaque PKS recombinante a conduit à la synthèse d'un produit. Ceci suggère donc que l'échange de domaine KR dans le module 2 n'affecte pas la fonctionnalité des PKS ainsi générées. L'échange de 9 KR différentes



de type A2 au niveau de cette PKS a permis la formation d'un produit épimérisé pour 7 d'entre elles. Celles-ci ont généré également le produit réduit mais non épimérisé. Les rendements les plus élevés en lactone épimérisée ont été dus aux échanges avec Amph KR<sub>1</sub> et Amph KR<sub>11</sub>, avec des rendements respectifs de 7 et 14 mg/L. Cependant l'échange de 7 KR de type B2 a abouti à la formation d'un produit épimérisé pour seulement 2 d'entre elles. À l'exception d'Elai KR<sub>4</sub> et de Con KR<sub>10</sub>, l'ensemble des constructions produisait la lactone non réduite en C3, la cétolactone. Ce polykétide est produit avec le rendement le plus élevé sauf pour Amph KR<sub>1</sub> et Amph KR<sub>11</sub>. Ce rendement variait de traces à 59 mg/L, alors que celui des lactones réduites variait de traces à 14 mg/L. Concernant les KR<sub>1</sub>, le rendement de chaque type de lactone produite était légèrement supérieur lorsque leur séquence nucléotidique a été insérée par le couple *BglIII/NheI* plutôt que *AvrII/NheI*, validant le choix de *BglIII* pour le deuxième cycle de clonage.

**VIP** Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide StereochemistryThibault Annaval,<sup>[a]</sup> Cédric Paris,<sup>[b]</sup> Peter F. Leadlay,<sup>[c]</sup> Christophe Jacob,<sup>\*[a]</sup> and Kira J. Weissman<sup>\*[a]</sup>

Modular polyketide synthases (PKSs) are multidomain multi-enzymes responsible for the biosynthesis in bacteria of a wide range of polyketide secondary metabolites of clinical value. The stereochemistry of these molecules is an attractive target for genetic engineering in attempts to produce analogues exhibiting novel therapeutic properties. The exchange of ketoreductase (KR) domains in model PKSs has been shown in several cases to predictably alter the configuration of the  $\beta$ -hydroxy functionalities but not of the  $\alpha$ -methyl groups. By sys-

tematic screening of a broad panel of KR domains, we have identified two donor KRs that afford modification of  $\alpha$ -methyl group stereochemistry. To the best of our knowledge, this provides the first direct in vivo evidence of KR-catalyzed epimerization. However, none of the introduced KRs afforded simultaneous alteration of methyl and hydroxy configurations in high yield. Therefore, swapping of whole modules might be necessary to achieve such changes in stereochemistry.

**Introduction**

Complex bacterial polyketides find wide use in clinical therapy, notably as antibiotics, immunosuppressants, and anti-cancer agents.<sup>[1]</sup> There is therefore significant interest in specific tailoring of their structures to generate analogues for evaluation as drug leads. A complementary approach to conventional chemical modification (semisynthesis) is to genetically engineer the underlying biosynthetic pathways.<sup>[2]</sup> This approach is particularly appealing for this type of polyketide, as the carbon skeletons are constructed on modular polyketide synthase (PKS) assembly lines in which each round of chain extension is handled by a specific set ("module") of enzymatic domains.<sup>[1]</sup> Minimally, the modules comprise an acyltransferase (AT) for building-block selection, a ketosynthase (KS) to join the blocks together, and an acyl carrier protein (ACP) domain, which serves as a tethering point for the growing chains. There can

be additional catalytic domains that chemically modify the newly formed intermediates: ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains. However, the promise of modular PKS synthetic biology remains hampered by our limited understanding of the structure and function of these multienzymes.<sup>[3]</sup>

Complex polyketides are typically rich in stereocenters, and the ability to alter the stereochemistry, especially for hydroxy groups that serve as attachment points for further enzymatic processing,<sup>[4]</sup> is a particularly desirable goal in PKS engineering. Our current model for stereocontrol by a modular PKS is shown in Scheme 1. The configuration of the hydroxy groups in a newly added chain-extension unit is established by the NADPH-dependent ketoreductase (KR) domains acting on the  $\beta$ -ketoacyl-ACP intermediates.<sup>[5]</sup> The face of the keto group to which the 4-*pro-S* hydride<sup>[6,7]</sup> is delivered is thought to be determined by the direction of substrate entry into the KR active site, thereby orienting it with respect to the NADPH cofactor.<sup>[8]</sup> Control of the 2-methyl configuration is more complicated, as it arises from the combined action of multiple domains—the AT responsible for selection of the methyl-branching extender unit methylmalonyl-CoA, the KS domains, which catalyze chain extension, and in many cases, the KR domains. Studies on 6-deoxyerythronolide B synthase (DEBS, the erythromycin PKS) revealed that all of the AT domains recruit the 2*S* and not the 2*R* isomer of methylmalonyl-CoA,<sup>[9,10]</sup> a choice that can be rationalized by inspecting the crystal structures of individual AT domains.<sup>[11]</sup> Based on homology among PKS ATs, it is assumed that all these domains exhibit the same stereochemical preference.

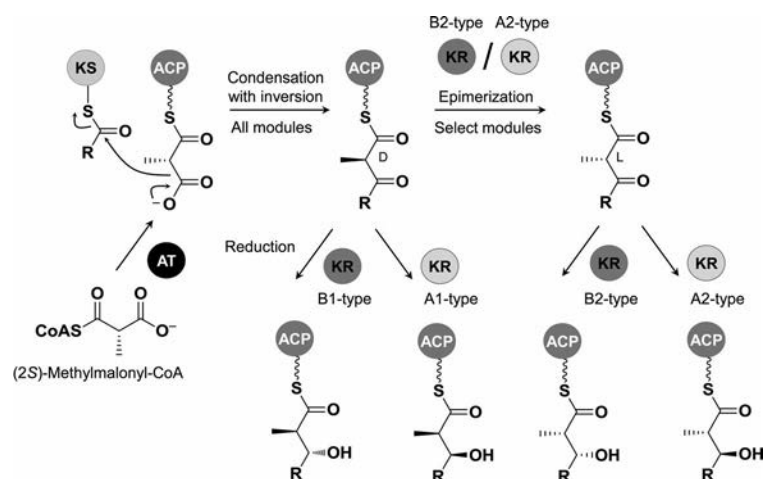
Condensation of the extender unit with the growing chain by the KS domains then proceeds with inversion of configuration at the methyl center,<sup>[12]</sup> thereby generating the *D*-config-

[a] T. Annaval, Dr. C. Jacob, Prof. K. J. Weissman  
UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA)  
CNRS–Université de Lorraine  
Biopôle de l'Université de Lorraine, Campus biologie-santé  
9 Avenue de la Forêt de Haye, CS 50184  
54505 Vandœuvre-lès-Nancy Cedex (France)  
E-mail: christophe.jacob@univ-lorraine.fr  
kira.weissman@univ-lorraine.fr

[b] C. Paris  
Laboratoire d'Ingénierie des Biomolécules  
Ecole Nationale Supérieure d'Agronomie  
et des Industries Alimentaires (ENSAIA)  
Université de Lorraine  
2, Avenue de la Forêt de Haye, B.P. 172  
54518 Vandœuvre-lès-Nancy Cedex (France)

[c] Prof. P. F. Leadlay  
Department of Biochemistry, University of Cambridge  
80 Tennis Court Road, Cambridge, CB2 1GA (UK)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201500113>.



**Scheme 1.** Unified model for stereocontrol in a modular PKS. The AT domains select a single isomer (2S) of the extender unit methylmalonate, and transfer it onto the ACP domain. KS-catalyzed Claisen-type condensation between the extender unit and the growing polyketide proceeds with inversion of stereochemistry at the methyl center and gives uniquely the D stereochemistry. The adjacent β-keto group can then be reduced in either direction by the KR domain to give an overall A1- or B1-type product; alternatively the methyl center can be epimerized by the KR domain prior to ketoreduction, thereby affording the A2- and B2-type configurations.

ured methyl in the resulting ACP-tethered intermediate. Although it was initially proposed that the KS domains were also responsible for epimerization thus yielding the L-configuration observed in many polyketides,<sup>[13,14]</sup> it is now clear that L-methyl groups arise from the epimerase activity of certain KR domains, even when the KR domains are inactive as ketoreductases.<sup>[15–18]</sup>

This understanding has led to the classification of modular PKS KR domains into six types:<sup>[15]</sup> those catalyzing ketoreduction but not epimerization (A1 and B1), ketoreduction and epimerization (A2 and B2), epimerization but not reduction (C2), and fully inactive KR domains (C1). Furthermore, specific combinations of KR sequence motifs (“fingerprints”) can be used to assign the domains to these categories, thus suggesting direct involvement of these residues in stereocontrol.<sup>[15]</sup> The direction of ketoreduction seems, therefore, to be an intrinsic property of the KR. In support of this, in the recently solved cryo-electron microscopy structure of a PKS module captured just prior to ketoreduction, the ACP domain carrying the polyketide intermediate docks against the KR, which positions the substrate to enter from either side of the active site.<sup>[19]</sup> However, although specific mutations of fingerprint residues altered the direction of reduction of model substrates *in vitro*,<sup>[20,21]</sup> the equivalent changes introduced into the PKS did not produce the same change in stereochemistry *in vivo*.<sup>[22]</sup> Additionally, extensive comparative analysis of these motifs does not reveal any obvious acidic or basic residues uniquely associated with epimerizing (A2/B2) KR domains.<sup>[17]</sup> Thus, the detailed molecular basis for stereocontrol by KR domains remains unresolved, hampering further attempts to alter stereochemistry by site-directed mutagenesis.

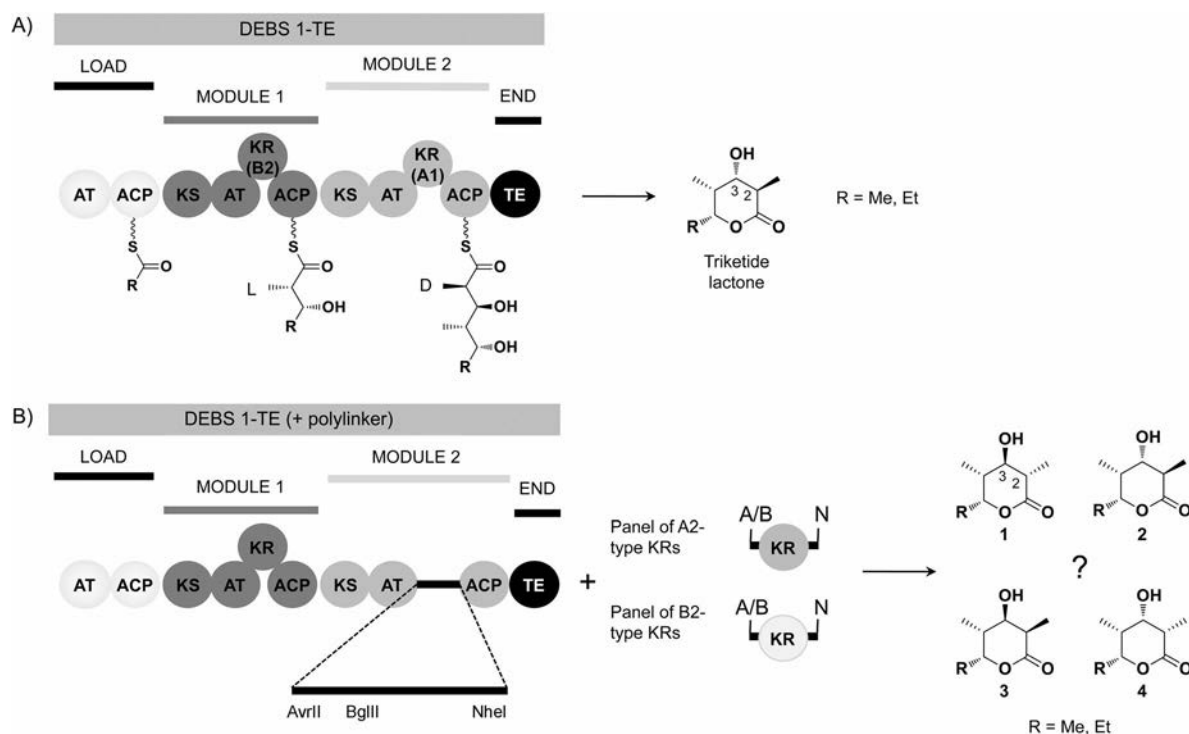
An alternative approach for manipulating the KR-catalyzed reaction involves exchanging one KR domain for another with

intrinsically opposite stereospecificity.<sup>[6]</sup> In this way, it is in principle possible to change the stereochemistry of the C2 methyl, the C3 hydroxy, or both simultaneously. This strategy has been previously explored by an approach termed “reductive loop swapping”<sup>[23]</sup> by using a model system derived from the DEBS PKS (DEBS 1-TE: the first multienzyme DEBS 1 fused to the chain-terminating thioesterase; Scheme 2). DEBS 1-TE catalyzes the release of the chain at the triketide stage, thereby giving a δ-lactone, which is highly amenable to structural analysis by gas chromatography–mass spectrometry (GC-MS). In this earlier work, the gene encoding KR<sub>2</sub> (the non-epimerizing A1-type KR) from the second module of DEBS 1-TE was initially replaced with a polylinker containing a number of restriction sites, thus allowing variation in the points of fusion between the introduced KR and the flanking AT and ACP domains of the parent PKS.<sup>[23]</sup> Inspection of the high-resolution structures of AT,<sup>[11]</sup> KR,<sup>[24]</sup> and ACP<sup>[25]</sup> domains confirmed that the chosen fusion points all lie outside of the conserved domains within the linker regions. By introducing genes encoding a range of B1-type (non-epimerizing) KR domains from DEBS and other PKSs into the polylinker, it

was possible to efficiently and predictably alter the stereochemistry of the ketoreduction. However, attempts to change the methyl stereochemistry (or both the methyl and hydroxy configurations together) by introducing one A2 and one B2 type KR, respectively, yielded only traces of the expected products.<sup>[23]</sup> Here, we examined whether the use of a more extensive pool of donor epimerizing KR domains improves the yield of methyl-epimerized products.

## Results and Discussion

In modules that incorporate only KR domains and no other reductive elements (DH or ER), KR epimerase activity can be directly detected by inspection of the stereochemistry of the corresponding methyl center in the final polyketide product (i.e., for a KR to be assigned as epimerizing, both its sequence and the product polyketide structure must be known). We therefore sourced KR domains from modules of this type, a choice which additionally simplified the definition of the domain boundaries (the boundaries are shown in Figure S1 in the Supporting Information). We obtained and cloned 14 such epimerizing KR domains:<sup>[26]</sup> eight A2-type KR domains (amphotericin (Amph) KR<sub>1</sub> and KR<sub>11</sub>,<sup>[27]</sup> concanamycin (Con) KR<sub>4</sub> and KR<sub>10</sub>,<sup>[28]</sup> elaiophyllin (Ela) KR<sub>4</sub>,<sup>[29]</sup> oligomycin (Olm) KR<sub>5</sub>,<sup>[30]</sup> pimarinic (PIMS) KR<sub>7</sub>,<sup>[31]</sup> and candidicin (Can) KR<sub>13</sub>)<sup>[32]</sup> and six B2-type KR domains (erythromycin (DEBS) KR<sub>1</sub>,<sup>[33]</sup> lankamycin (Lkm) KR<sub>1</sub>,<sup>[34]</sup> pikromycin (Pik) KR<sub>1</sub>,<sup>[35]</sup> lasalocid (Las) KR<sub>7</sub>,<sup>[36]</sup> ECO-02301 (ECO) KR<sub>19</sub>,<sup>[37]</sup> and stambomycin (Sam) KR<sub>21</sub>).<sup>[38]</sup> For KR domains that are preceded by a recently identified dimerization domain,<sup>[39]</sup> the sequence encoding this region was also included, as done previously.<sup>[23]</sup> The fact that this pool of KR domains operates on chains of vastly different length (from five-carbon chains in the case of KR<sub>1</sub> to a 45-carbon chain for Sam KR<sub>21</sub>) allowed us to explore potential



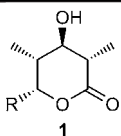
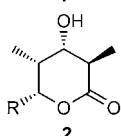
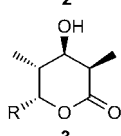
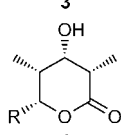
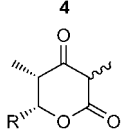
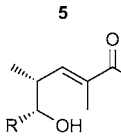
**Scheme 2.** Model biosynthetic system. A) The parent DEBS 1-TE was generated by relocating the C-terminal thioesterase of DEBS 3 to the end of DEBS 1. The resulting mini-PKS produces a triketide lactone from either acetate or propionate as the starter unit. B) Derivative of DEBS 1-TE amenable to exchange of KR domains, created by replacing the A1-type KR domain (KR<sub>2</sub>) with a polylinker. In this work, various A2- and B2-type KR domains were introduced into the polylinker at BglIII/NheI or AvrII/NheI restriction sites. The resulting triketide lactones were then analyzed by GC-MS to determine the stereochemistry at positions 2 and 3.

links between native substrate size and control of stereochemistry during the reduction of a model triketide.

We also cloned the two KR from the mycolactone (MIs) PKS: KR<sub>A</sub>, which is present in identical copies in three modules, and KR<sub>B</sub>, which occurs in 13 modules.<sup>[40]</sup> In its native contexts, KR<sub>A</sub> reduces intermediates lacking an  $\alpha$ -methyl, but the domain can act on  $\alpha$ -methyl-containing substrates (at least in vitro), with a preference for those containing an L configuration.<sup>[41]</sup> In all but one of the modules in which it is active, KR<sub>B</sub> does not appear to catalyze methyl-group epimerization; however, the product of module 2 contains an L-methyl group, thus suggesting that KR<sub>B</sub> epimerizes this center.<sup>[40]</sup> Given these observations, it was of interest to see how both these KR would perform in a non-native context against substrates bearing an  $\alpha$ -D-methyl. A detailed comparison of the sequences of KR<sub>B</sub> from module 2 with that from module 1 (a putatively non-epimerizing KR<sub>B</sub>) revealed three residues that differ,<sup>[42]</sup> although they do not account convincingly for why the module 2 KR should be epimerizing (Figure S2). Nonetheless, we introduced these three residues into our initially cloned KR-coding sequence by site-directed mutagenesis, and simultaneously removed the internal BglIII and NheI restriction sites. By comparative sequence analysis (Figure S1), it was clear that the linker region into which the 5' splice site should be inserted was upstream of the DH domain preceding KR<sub>B</sub>, and therefore we amplified DH<sub>2</sub> along with KR<sub>B</sub>. Sequence analysis (Figure S1) also suggested that this DH domain should be active, although the

product structure shows that it is not. Given this ambiguity, it was unclear as to whether DH<sub>2</sub> would be functional in its new context.

Each KR was introduced into the previously described DEBS 1-TE model system<sup>[23]</sup> by using 5' and 3' polylinker restriction sites (BglIII and NheI, respectively; Figure S1); this reproducibly gave high yields in the initial experiments with a variety of KR donors. In the case of the KR from module 1, we additionally used a 5' AvrII site, as these KR were previously reported not to tolerate substrates larger than diketides,<sup>[16]</sup> and so we thought it prudent to choose more than one fusion point in the target PKS. To facilitate heterologous expression in *Saccharopolyspora erythraea* JC2,<sup>[43]</sup> the original DEBS 1-TE-encoding gene was transferred from plasmid pJLK117<sup>[23]</sup> to plasmid pMU3 (*Escherichia coli*/*Streptomyces* shuttle vector), which contains *oriT* and an *attP* integration site from bacteriophage VWB. The KR-modified DEBS 1-TE constructs in pMU3 were transformed into *S. erythraea* JC2 by transconjugation with *E. coli* ET12567(pUZ8002). JC2 is a derivative of the native erythromycin-producing strain that lacks the DEBS genes (*eryA*) except the region encoding the terminal TE domain.<sup>[43]</sup> Thus, in this strain, integration of pMU3-based plasmids encoding DEBS 1-TE can occur either at the *attB* site or at the TE gene. Previous experiments have shown that lactone yields for these two sites are essentially the same.<sup>[44]</sup> Following selection of transformants grown as colony "patches", identical-size samples of the transconjugants (at least five clones per plasmid)

Catalytic events	Product	$t_R$ [min]	
		R = Et (a)	R = Me (b)
epimerization/B-type reduction		15.2	n.d.
no epimerization/A-type reduction		15.7	14.7
no epimerization/B-type reduction		15.9	14.9
epimerization/A-type reduction		16.1	15.1
no reduction		14.3	13.3
epimerization/B-type reduction/ dehydration		14.1 <sup>[a]</sup>	12.4 <sup>[a]</sup>

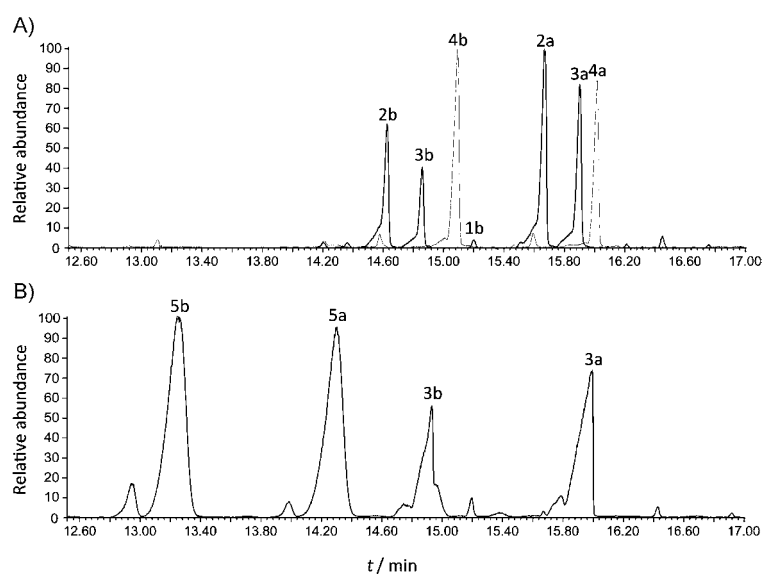
[a] Analysis by UHPLC-MS following methylation; all others by GC-MS.  
n.d.: not detected.

were grown in liquid TSB medium for two days at 30 °C. Equal portions of the cultures were then used to inoculate SM3 production plates, which were incubated for 14 days at 30 °C. Equal-size portions of each SM3 plate were removed and analyzed.

In *S. erythraea* JC2, the mini-PKS uses propionate (**a** forms, Table 1) and acetate (**b** forms) as the starter unit (Scheme 1, R = Me or Et; a:b ratios in JC2 between 1:1.6 and 2.6:1). Assignment of product identity and stereochemistry by GC-MS on an Agilent 7890-5975C was based on the availability of three of the triketide lactone diastereoisomers at positions 2 and 3 (**1–3**, Table 1) and the 3-ketolactone **5**; **4** was assigned by elimination, as only four reduced products are possible (positions 4 and 5 are determined by module 1, which was not altered in this experiment). Lactones **1 a** and **2 a** were derived by chemical synthesis<sup>[17,45,46]</sup> and are additionally generated by strains JC2/pJLK30 and JC2/pJLK25 along with **1 b** and **2 b**.<sup>[23]</sup> **2 a** was a kind gift from A. Garg and D.

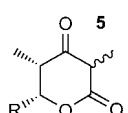
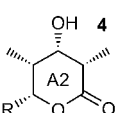
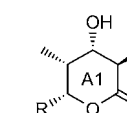
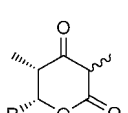
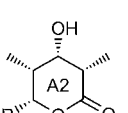
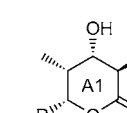
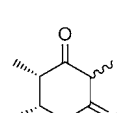
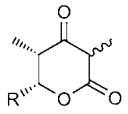
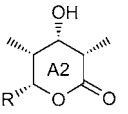
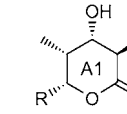
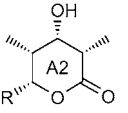
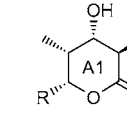
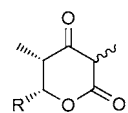
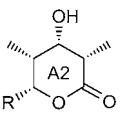
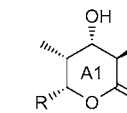
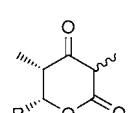
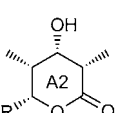
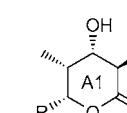
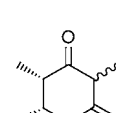
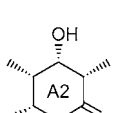
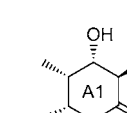
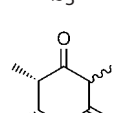
Cane, Brown University, USA); lactones **3 a**, **3 b**, and ketolactones **5 a** and **5 b** are previously identified products of strains JC2/pJLK30 and JC2/pJLK35.<sup>[23]</sup> Strains JC2/pJLK25, JC2/pJLK30, and JC2/pJLK35 were therefore grown alongside the newly engineered strains to serve as controls. The lactone products **1–5** were well resolved, gave similar retention times as observed previously (Figure 1 and Table 1),<sup>[23,45]</sup> and exhibited masses and fragmentation patterns consistent with their structures (Figure S3).<sup>[47]</sup> To quantify product yields, we generated a standard curve from a sample of **2 a**. This showed an essentially linear increase in peak area in the resulting chromatograms for concentrations of product between 6 and 100 mg L<sup>-1</sup> (Figure S4), thus allowing determination of product yields in this range either by direct analysis of the biosynthetic sample or after suitable dilution (yields below 6 mg L<sup>-1</sup> are therefore estimates). Two clones of each construct were analyzed (Table 2; a representative analysis of each construct is provided in Figure S5).

All KR domains were introduced at the same two restriction sites (BglII and NheI) within the polylinker. The fact that all but one (Ela KR<sub>4</sub>) of the resulting multienzymes produced at least one lactone in yields comparable to the earlier swap experiments (traces–59 mg L<sup>-1</sup> vs. 5.5–27 mg L<sup>-1</sup>)<sup>[23]</sup> validated this choice, and demonstrated that all these engineered PKSs were functional. Nonetheless, the yields were uniformly lower than those obtained from the parent DEBS 1-TE multienzyme (100 mg L<sup>-1</sup>), thus showing that KR swaps compromise enzyme activity.<sup>[23]</sup> Use of the alternative 5' AvrII site to insert the KR<sub>1,5</sub> uniformly lowered yields, thus again showing that splice site selection is a crucial parameter in these experiments. For all nine constructs in which the introduced KR<sub>s</sub> were active for reduction, only the correct sense of hydride delivery was observed (A-type for A2 KR<sub>s</sub>, B-type for B2 KR<sub>s</sub>). These data are



**Figure 1.** GC-MS identification of the lactones. Standards for lactones **1–3** and **5** were obtained from previously investigated strains.<sup>[23]</sup> A) **2 a**, **2 b** from JC2/pJLK25; **1 a**, **3 a**, **3 b** from JC2/pJLK30; B) **5 a**, **5 b** from JC2/pJLK35. Lactones **1 a** and **2 a** were verified by comparison to synthetic standards. The identities of lactones **4 a** and **4 b** (shown here, JC2/Amph KR<sub>1</sub>) were assigned by elimination.

**Table 2.** Results of introduction of A2- and B2-type KRs into DEBS 1-TE.

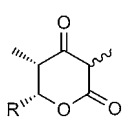
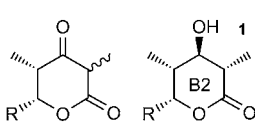
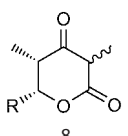
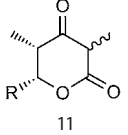
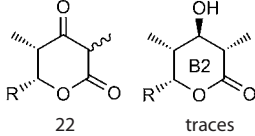
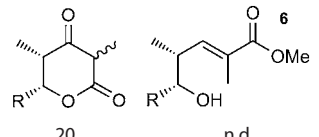
Introduced KR	Type	Chain length of native substrate	Products and yields [mg L <sup>-1</sup> ] <sup>a</sup>		
Amph KR <sub>1</sub>	A2	C <sub>4</sub>	 <b>5</b> B/N: 6 A/N: 5	 <b>4</b> A2 B/N: 7 A/N: ~0.5	 <b>2</b> A1 B/N: traces A/N: traces
Con KR <sub>4</sub>	A2	C <sub>10</sub>	 <b>29</b>	 A2 traces	 A1 traces
Ela KR <sub>4</sub>	A2	C <sub>10</sub>	no product		
Olm KR <sub>5</sub>	A2	C <sub>12</sub>	 <b>52</b>		
PIMS KR <sub>7</sub>	A2	C <sub>16</sub>	 <b>18</b>	 A2 ~2	 A1 traces
Con KR <sub>10</sub>	A2	C <sub>22</sub>	 A2 traces		 A1 traces
Amph KR <sub>11</sub>	A2	C <sub>24</sub>	 ~3	 A2 <b>14</b>	 A1 traces
Can KR <sub>13</sub>	A2	C <sub>26</sub> + amino benzoyl	 ~5	 A2 ~3	 A1 traces
Mls KR <sub>A</sub>	A	C <sub>4</sub> , C <sub>6</sub> , C <sub>12</sub>	 ~3	 A2 traces	 A1 traces
DEBS KR <sub>1</sub>	B2	C <sub>5</sub>	 B/N: 59 A/N: 6		

consistent with the idea that the direction of ketoreduction is a fundamental property of the KR active site,<sup>[5,19]</sup> thus explaining why the majority of KRs tested were shown to generate the anticipated hydroxy configurations in new contexts.<sup>[5,14]</sup>

In total, nine of the 16 KR exchanges resulted in product with a C2-epimerized methyl group, consistent with the intrinsic epimerase activity of this panel of KRs. Yields ranged from trace amounts up to 7 and 14 mg L<sup>-1</sup> (Amph A2-type KRs 1 and 11 respectively; Table 2), a significant improvement over previous attempts.<sup>[23]</sup> These data also demonstrate that the TE domain can, as predicted, efficiently release diastereomeric product from the engineered multi-enzyme.<sup>[48]</sup> Nonetheless, that nearly half of the donor KRs did not switch the product configuration shows that a detailed understanding of the molecular basis for epimerization remains elusive. Interestingly, Mls KR<sub>A</sub>, which normally reduces a C<sub>α</sub>-methylene product, was found to epimerize an adjacent methyl center, albeit inefficiently.

The B2-type Lkm KR<sub>1</sub> and Sam KR<sub>21</sub> only performed ketoreduction when the adjacent methyl center was epimerized, and gave exclusively **1**. In contrast, all the A2-type KRs produced trace quantities of **2** incorporating a non-epimerized methyl group in addition to **4**, in which the methyl is epimerized, thus showing that epimerization is not an absolute prerequisite for ketoreduction by these KRs. As A2-type KRs normally generate a single (epimerized) product, these data argue that the context in which the KR operates can have an effect on the stereochemical outcome. This observation hints that the mechanism of epimerization depends not only on the KR but additionally on an inherent property of the host module (e.g., providing water access to the KR active site). Indeed, it has been observed previously that the DEBS module 2 studied here is substantially more resistant to proteolysis than the upstream module 1, in which programmed epimerization occurs, thus implying that module 2 adopts a more compact architecture.<sup>[49]</sup> This factor may further complicate attempts to rationally introduce epimerized methyl groups into specific positions of designed polyketides by domain swapping.

In three cases of introduced A2-type KRs (Amph KR<sub>1</sub> (BgIII/Nhel), Amph KR<sub>11</sub>, and Con KR<sub>10</sub>), the epimerized lactone **4** was the major product. In contrast, in the remaining six experiments in which reduced lactones were produced, the predominant product was the unreduced ketolactone (Table 2), thus demonstrating that these introduced A2- and B2-type KRs operated inefficiently in this context relative to their A1- and B1-type counterparts (yields 25–100 mg L<sup>-1</sup>).<sup>[23]</sup> No obvious correlation was seen between the efficiency of ketoreduction and chain length of the native substrate of each KR, as Amph

Table 2. (Continued)			
Introduced KR	Type	Chain length of native substrate	Products and yields [mg L <sup>-1</sup> ] <sup>[a]</sup>
Pik KR <sub>1</sub>	B2	C <sub>5</sub>	 B/N: 37 A/N: 22
Lkm KR <sub>1</sub>	B2	C <sub>6</sub>	 B/N: 6 A/N: ~2 B/N: traces A/N: traces
Las KR <sub>7</sub>	B2	C <sub>16</sub>	 8
ECO KR <sub>19</sub>	B2	C <sub>42</sub> + NH <sub>2</sub>	 11
Sam KR <sub>21</sub>	B2	C <sub>45</sub>	 22 traces
Mls KR <sub>B</sub>	B1/ B2	C <sub>4r</sub> , C <sub>6r</sub> , C <sub>8r</sub> , C <sub>10r</sub> , C <sub>12r</sub> , C <sub>14r</sub> , C <sub>16r</sub> , C <sub>18r</sub> , C <sub>20</sub>	 20 n.d. 6

[a] Combined yield of **a** and **b** forms. n.d.: not detected.

KRs 1 and 11 gave the highest yields of reduced relative to ketolactone products.

For the remaining six constructs (A-type KR Olm KR<sub>5</sub> and B-type KRs DEBS KR<sub>1</sub>, Pik KR<sub>1</sub>, Las KR<sub>7</sub>, ECO KR<sub>19</sub>, and Mls KR<sub>B</sub>) only the ketolactone product **5** was observed, thus showing that although the overall multienzyme was operational, the KRs did not work in their new positions. It has previously been suggested that KR<sub>1</sub> domains are poor choices for exchange into later modules as they cannot tolerate substrates longer than their native diketides.<sup>[45]</sup> Although this idea tallies with some of the results obtained here, the substrates of Olm KR<sub>5</sub>, Las KR<sub>7</sub>, ECO KR<sub>19</sub>, and Mls KR<sub>B</sub> are substantially longer, and Amph and Lkm PKS KR<sub>s</sub> were both functional in their new contexts (Table 2). That Mls KR<sub>B</sub> failed to act was particularly surprising, given its broad substrate tolerance and the fact that it can function as both a B1- and B2-type KR. To address the

possibility that the DH upstream of the introduced Mls KR<sub>B</sub> was in fact active and so catalyzing dehydration of the initially formed hydroxy group to the predicted *trans* double bond, we methylated the initial extracts of the strain with trimethylsilyl diazomethane (TMS-diazomethane). We then reanalyzed for the presence of the predicted open chain methyl ester **6** (Table 1) by comparison with extracts of the previously described **6**-producing strain, JC2/pJLK29.<sup>[23]</sup> **6** was present in extracts of JC2/pJLK29 (Figure S5), but was not detected in extracts of Mls KR<sub>B</sub>. We could therefore conclude that Mls KR<sub>B</sub> was indeed non-functional in its new context. Taken together, these experiments revealed that five of the six inactive KRs were B-type, and suggest that these KRs are inherently less amenable to relocation (at least into an A1 context) than their A-type counterparts.

## Conclusion

In previous experiments, exchange of KR domains has proved a robust approach for predictably altering the stereochemistry of ketoreduction within a target module.<sup>[5,23]</sup> Out of a total of eight screened KR domains, we have identified two A2-type KR domains (Amph KRs 1 and 11) that, when introduced through validated splice sites into an A1-modular context, afford modification of methyl group stereochemistry at moderate to good efficiency. Given this hit rate, it seems likely that further A2-type KRs amenable to such relocation will be identified by ongoing genome-sequencing efforts. In contrast, the Amph KRs produced mixtures that included non-epimerized lactone and substantial quantities of unreduced keto-lactone. Furthermore, B2-type KRs, which would permit simultaneous alteration of both methyl and hydroxy configurations, appear to have generally low portability into an A1 modular context. Taken together, these results suggest that to obtain both efficient ketoreduction and a change in methyl group stereochemistry will require swapping complete modules between PKS systems instead of just the responsible KR domains.

## Experimental Section

**Materials, DNA isolation, and DNA manipulation:** *Streptomyces* strains, cosmids, BACs and plasmids used in this study to source the DNA encoding the introduced KR domains are provided in Table S1. Standard procedures for DNA isolation and manipulation were performed.<sup>[50,51]</sup> Isolation of DNA fragments from agarose gel and purification of PCR products were carried out with a NucleoSpin Extract II kit (Macherey–Nagel, Hoerd, France). Standard PCR reactions were performed with Phusion high-fidelity DNA polymerase (Thermo Scientific) in a Mastercycler Pro (Eppendorf, Wesseling-Berzdorf, Germany). Synthetic oligonucleotides were purchased from Sigma–Aldrich, and DNA sequencing was carried out by GATC Biotech (Mulhouse, France). The three residues that differ

between the two KR<sub>B</sub>s from modules 1 and 2 of the mycolactone PKS were mutated by standard PCR site-directed mutagenesis: A5459, Q5515, and H5550 (numbering according to GenBank CAE46851) were mutated to V, L, and R respectively (oligos in Table S2). In addition, single internal BglIII and NheI restriction sites in KR<sub>B</sub> were eliminated by standard PCR site-directed mutagenesis without introducing changes at the amino acid level (Table S2). All mutations were confirmed by sequencing.

**Bacterial strains and culture conditions:** *E. coli* DH5 $\alpha$  was used in all standard cloning procedures and was grown in LB medium; *E. coli* ET12567 (pUZ8002)<sup>[52]</sup> was used in transconjugation experiments with *S. erythraea* JC2. *S. erythraea* mutant strain JC2 NCIMB 40802 (in which the *eryA* genes are deleted, apart from the thioesterase domain) has been described previously.<sup>[44]</sup> ABB13 growth medium: bacto soytone (5 g), soluble starch (5 g), CaCO<sub>3</sub> (3 g), MOPS (2.1 g), and agar (20 g) in 1 L (pH 7.2). After autoclaving, thiamine (1%, 1 mL), FeSO<sub>4</sub>·7H<sub>2</sub>O (1.2%, 1 mL) and MgCl<sub>2</sub> (10 mM) were added. Screening medium 3 (SM3): glucose (4 g), MD30 (40 g), soya flour (20 g), beet molasses (2.4 g), KH<sub>2</sub>PO<sub>4</sub> (0.20 g), and CaCO<sub>3</sub> (2 g), in 1 L (pH 7.0).

**Construction of pMU3-DEBS 1-TE expression plasmids containing alternative KR domains:** The DEBS 1-TE-encoding insert from pJLK117<sup>[23]</sup> was excised as an NdeI/XbaI fragment and ligated into NdeI/XbaI-digested pMU3<sup>[43]</sup> to give pMU3-DEBS 1-TE. Each KR domain was amplified from cosmid or genomic template by using appropriate pairs of synthetic oligonucleotide primers (Table S1) that introduced flanking restriction sites for direct cloning into pMU3-DEBS 1-TE. KR domains from module 1 were amplified as both AvrII/NheI and BglIII/NheI fragments; the remaining KR domains were amplified only as BglIII/NheI fragments.

**Transformation of *S. erythraea* JC2 by transconjugation, and growth for triketide lactone production:** The 19 expression plasmids were used to transform *E. coli* ET12567 (pUZ8002). The plasmids were then introduced into *S. erythraea* by transconjugation.<sup>[43]</sup> Transconjugants were selected by resistance to apramycin (acquired upon integration of the plasmid into the *S. erythraea* chromosome). After 13 days, single transformants were picked and grown as patches for 13 days on ABB13 plates supplemented with apramycin and naladixic acid. Identically sized samples of the transconjugants (at least five clones per plasmid) were then grown in liquid TSB for two days at 30 °C. Equal portions of the cultures were then used to inoculate SM3 production plates, which were incubated for 14 days at 30 °C. Equally sized portions of each SM3 plate (2.2 cm<sup>2</sup>) were removed and frozen so that all samples could be analyzed together.

**Extraction of triketide lactones:** Extraction of triketides was done by the "lactonex" procedure: the excised 2.2 cm<sup>2</sup> pieces of mycelium were chopped into pieces and transferred into 2 mL Eppendorf tubes. Ethyl acetate (1.2 mL) and formic acid (20  $\mu$ L) were added with mixing. The tubes were incubated at 50 °C for 15 min, and then vortexed for 30 min. The mixtures were centrifuged (20 000 g, 1 min), and the supernatants were placed into 1.5 mL Eppendorf tubes. After evaporation to dryness under reduced pressure, the residue in each tube was dissolved in ethyl acetate (50  $\mu$ L), centrifuged (20 000 g, 1 min), and the supernatant was transferred to a GC-MS vial for analysis.

**Analysis of triketide fermentation products by GC-MS and LC-MS:** GC-MS was performed on a model 7890A-5975C system (Agilent Technologies) equipped with a J&W DB 5 MS column (30 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m; Agilent Technologies) in EI mode: 40 °C for 2 min; 10 °C min<sup>-1</sup> to 250 °C; 25 °C min<sup>-1</sup> to 300 °C. The quantities

of all lactones present in each sample were determined by integration of the respective peaks and comparison to the standard curve (Figure S4). The identities of lactones 1–3 and 5 were confirmed in each case by comparison of the observed retention times and mass spectra with those of synthetic<sup>[45,46]</sup> and/or biosynthetic standards.<sup>[23]</sup> The identity of 4 was assigned by elimination. Following methylation of extracts of JC2/pJLK29, MIs KR<sub>B</sub>, and MIs KR<sub>A</sub> (negative control) with TMS-diazomethane, LC-MS was performed on a Dionex UltiMate 3000-LTQ UHPLC-MS system (Thermo Scientific) equipped with an Alltech Alltima C18 column (150  $\times$  2.1 mm; 5  $\mu$ m) at 25 °C (flow-rate, 0.2 mL min<sup>-1</sup>). Metabolites were eluted over 20 min with a linear gradient: 5–95% phase B (acetonitrile containing HCOOH (0.1%, v/v)) in phase A (water containing HCOOH (0.1%, v/v)). The products were analyzed by ESI-MS in positive-ion MS mode, in both the full-scan MS mode and by MS<sup>2</sup> of *m/z* 187.

## Acknowledgements

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the University of Lorraine (UL), and the Conseil Régional de Lorraine (BQR grant to K.J.W.). P.F.L. is an International Research Awardee of the Alexander von Humboldt Foundation. We would like to thank the following for provision of DNA or bacterial strains: Patrick Caffrey (*Streptomyces nodusus*), Kenji Arakawa (cosmid 14F1), José A. Gil (*S. griseus*), Yeo J. Yoon (*S. venezuelae*), Jesús F. Aparicio (*S. natalensis*), and Bertrand Aigle (genomic DNA of *S. ambofaciens*). Arnaud Gruez is thanked for help with preparing figures, and David A. Cane and Ashish Garg are thanked for providing lactone standard 2a.

**Keywords:** biosynthesis • genetic engineering • ketoreductases • polyketides • stereocontrol

- [1] C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688–4716; *Angew. Chem.* **2009**, *121*, 4782–4811.
- [2] K. J. Weissman, P. F. Leadlay, *Nat. Rev. Microbiol.* **2005**, *3*, 925–936.
- [3] S. Poust, A. Hagen, L. Katz, J. D. Keasling, *Curr. Opin. Biotechnol.* **2014**, *30*, 32–39.
- [4] F. Schlünzen, R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath, F. Franceschi, *Nature* **2001**, *413*, 814–821.
- [5] C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **1998**, *120*, 2478–2479.
- [6] M. McPherson, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **1998**, *120*, 3267–3268.
- [7] Y. Yin, R. Gokhale, C. Khosla, D. E. Cane, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1477–1479.
- [8] A. P. Siskos, A. Baerga-Ortiz, S. Bali, V. Stein, H. Mamdami, D. Spittler, B. Popovic, J. B. Spencer, J. Staunton, K. J. Weissman, P. F. Leadlay, *Chem. Biol.* **2005**, *12*, 1145–1153.
- [9] A. F. Marsden, P. Caffrey, J. F. Aparicio, M. S. Loughran, J. Staunton, P. F. Leadlay, *Science* **1994**, *263*, 378–380.
- [10] K. E. H. Wiesmann, J. Cortés, M. J. B. Brown, A. L. Cutter, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1995**, *2*, 583–589.
- [11] Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11124–11129.
- [12] K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton, P. F. Leadlay, *Biochemistry* **1997**, *36*, 13849–13855.
- [13] I. Böhm, I. E. Holzbaur, U. Hanefeld, J. Cortés, J. Staunton, P. F. Leadlay, *Chem. Biol.* **1998**, *5*, 407–412.
- [14] I. E. Holzbaur, A. Ranganathan, I. P. Thomas, D. J. A. Kearney, J. A. Reather, B. A. M. Rudd, J. Staunton, P. F. Leadlay, *Chem. Biol.* **2001**, *8*, 329–340.



- [15] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.
- [16] C. R. Valenzano, R. J. Lawson, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2009**, *131*, 18501–18511.
- [17] A. Garg, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2013**, *135*, 16324–16327.
- [18] A. Garg, X. Xie, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2014**, *136*, 10190–10193.
- [19] J. R. Whicher, S. Dutta, D. A. Hansen, W. A. Hale, J. A. Chemler, A. M. Dosey, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith, G. Skiniotis, *Nature* **2014**, *510*, 560–564.
- [20] A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spittler, M. G. Williams, N. Campillo, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 277–285.
- [21] H. M. O'Hare, A. Baerga-Ortiz, B. Popovic, J. B. Spencer, P. F. Leadlay, *Chem. Biol.* **2006**, *13*, 287–296.
- [22] D. H. Kwan, M. Tosin, N. Schläger, F. Schulz, P. F. Leadlay, *Org. Biomol. Chem.* **2011**, *9*, 2053–2056.
- [23] L. Kellenberger, I. S. Galloway, G. Sauter, G. Böhm, U. Hanefeld, J. Cortés, J. Staunton, P. F. Leadlay, *ChemBioChem* **2008**, *9*, 2740–2749.
- [24] A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737–748.
- [25] V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi, C. Khosla, *Protein Sci.* **2007**, *16*, 2093–2107.
- [26] A. Kitsche, M. Kalesse, *ChemBioChem* **2013**, *14*, 851–861.
- [27] P. Caffrey, S. Lynch, E. Flood, S. Finnan, M. Oliynyk, *Chem. Biol.* **2001**, *8*, 713–723.
- [28] S. F. Haydock, A. N. Appleyard, T. Mironenko, J. Lester, N. Scott, P. F. Leadlay, *Microbiology* **2005**, *151*, 3161–3169.
- [29] S. F. Haydock, T. Mironenko, H. I. Ghoorahoo, P. F. Leadlay, *J. Biotechnol.* **2004**, *113*, 55–68.
- [30] S. Omura, H. Ikeda, J. Ishikawa, A. Hanamoto, C. Takahashi, M. Shinose, Y. Takahashi, H. Horikawa, H. Nakazawa, T. Osonoe, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12215–12220.
- [31] J. F. Aparicio, R. Fouces, M. V. Mendes, N. Olivera, J. F. Martín, *Chem. Biol.* **2000**, *7*, 895–905.
- [32] S. Chen, X. Huang, X. Zhou, L. Bai, J. He, K. J. Jeong, S. Y. Lee, Z. Deng, *Chem. Biol.* **2003**, *10*, 1065–1076.
- [33] J. Cortés, S. F. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **1990**, *348*, 176–178.
- [34] K. Arakawa, F. Sugino, K. Kodama, T. Ishii, H. Kinashi, *Chem. Biol.* **2005**, *12*, 249–256.
- [35] Y. Xue, L. Zhao, H.-W. Liu, D. H. Sherman, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12111–12116.
- [36] L. Smith, H. Hong, J. B. Spencer, P. F. Leadlay, *ChemBioChem* **2008**, *9*, 2967–2975.
- [37] J. B. McAlpine, B. O. Bachmann, M. Pirae, S. Tremblay, A.-M. Alarco, E. Zazopoulos, C. M. Farnet, *J. Nat. Prod.* **2005**, *68*, 493–496.
- [38] L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis, B. Aigle, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6258–6263.
- [39] J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1263–1270.
- [40] T. P. Stinear, A. Mve-Obiang, P. L. C. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. R. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay, S. T. Cole, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1345–1349.
- [41] S. Bali, K. J. Weissman, *ChemBioChem* **2006**, *7*, 1935–1942.
- [42] A. T. Keatinge-Clay, *Nat. Prod. Rep.* **2012**, *29*, 1050–1073.
- [43] C. J. Rowe, J. Cortés, S. Gaisser, J. Staunton, P. F. Leadlay, *Gene* **1998**, *216*, 215–223.
- [44] K. J. Weissman, *ChemBioChem* **2006**, *7*, 485–494.
- [45] R. Castonguay, W. He, A. Y. Chen, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2007**, *129*, 13758–13769.
- [46] A. L. Wilkinson (née Cutter), U. Hanefeld, B. Wilkinson, P. F. Leadlay, J. Staunton, *Tetrahedron Lett.* **1998**, *39*, 9827–9830.
- [47] K. J. Weissman, G. C. Kearney, P. F. Leadlay, J. Staunton, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2103–2108.
- [48] K. J. Weissman, C. J. Smith, U. Hanefeld, R. Aggarwal, M. Bycroft, J. Staunton, P. F. Leadlay, *Angew. Chem. Int. Ed.* **1998**, *37*, 1437–1440; *Angew. Chem.* **1998**, *110*, 1503–1506.
- [49] H. Hong, A. N. Appleyard, A. P. Siskos, J. Garcia-Bernardo, J. Staunton, P. F. Leadlay, *FEBS J.* **2005**, *272*, 2373–2387.
- [50] J. Sambrook, D. Russell, *Molecular Cloning: A Laboratory Manual* (3rd ed.) Cold Spring Harbor Laboratory Press, New York, **2000**.
- [51] T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, UK, **2000**.
- [52] P. Mazodier, R. Petter, C. Thompson, *J. Bacteriol.* **1989**, *171*, 3583–3585.

Manuscript received: March 5, 2015

Accepted article published: April 7, 2015

Final article published: May 28, 2015

# CHEMBIOCHEM

## Supporting Information

### **Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide Stereochemistry**

Thibault Annaval,<sup>[a]</sup> Cédric Paris,<sup>[b]</sup> Peter F. Leadlay,<sup>[c]</sup> Christophe Jacob,<sup>\*[a]</sup> and Kira J. Weissman<sup>\*[a]</sup>

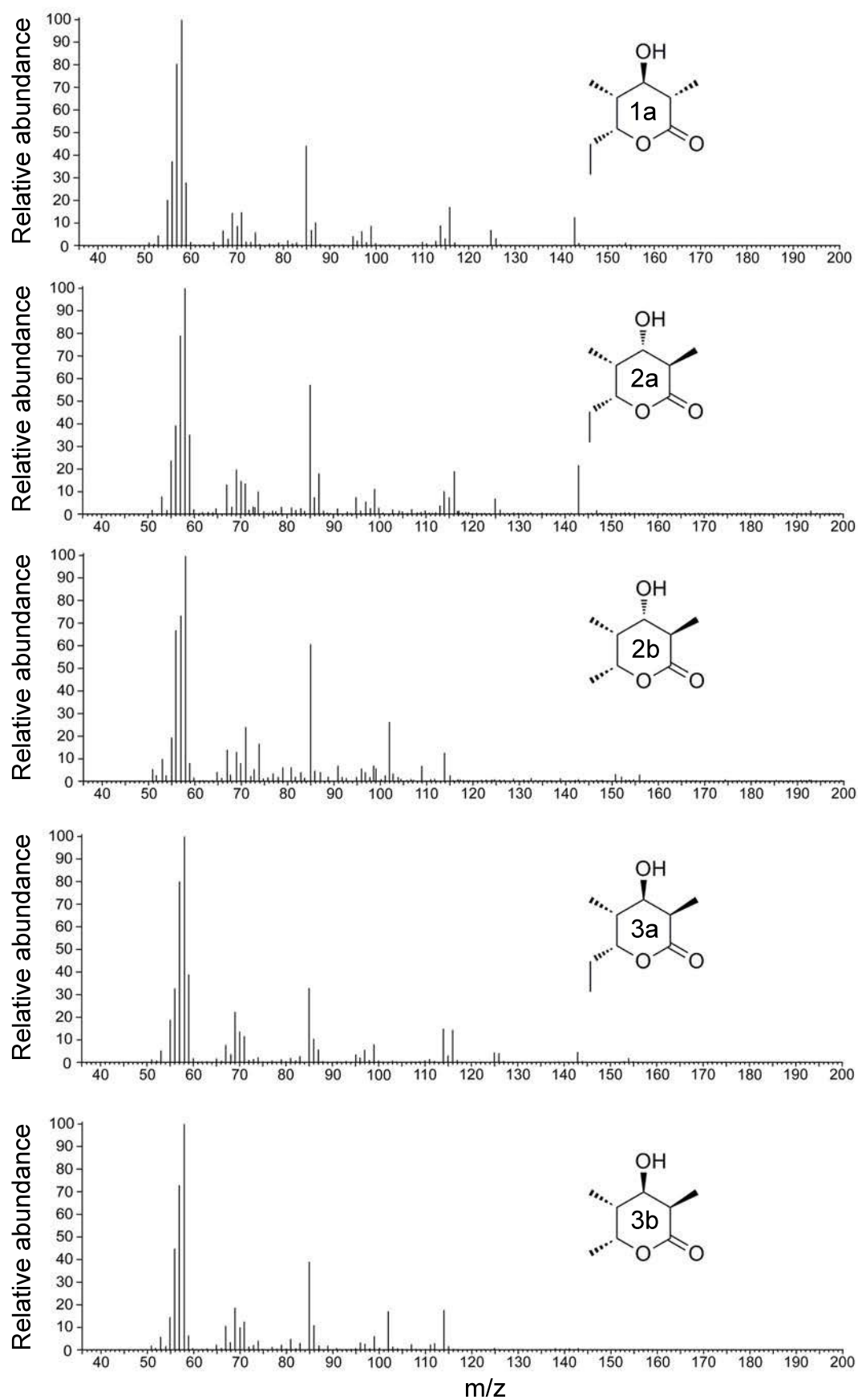
cbic\_201500113\_sm\_miscellaneous\_information.pdf

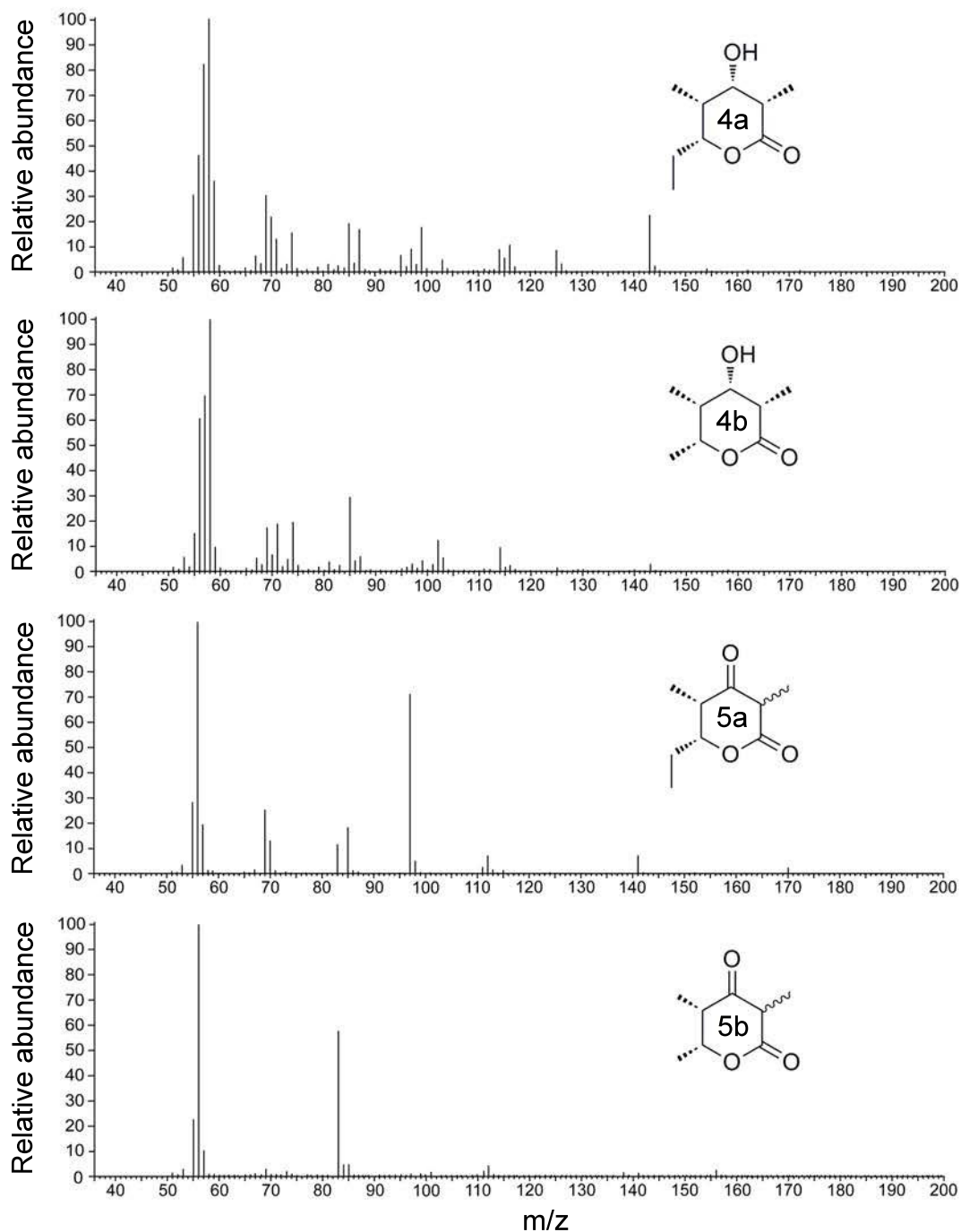






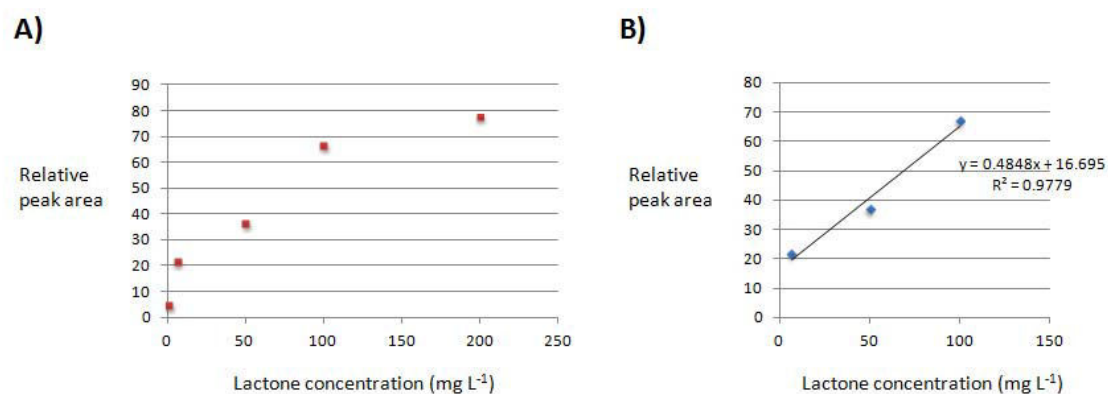
Figure S3





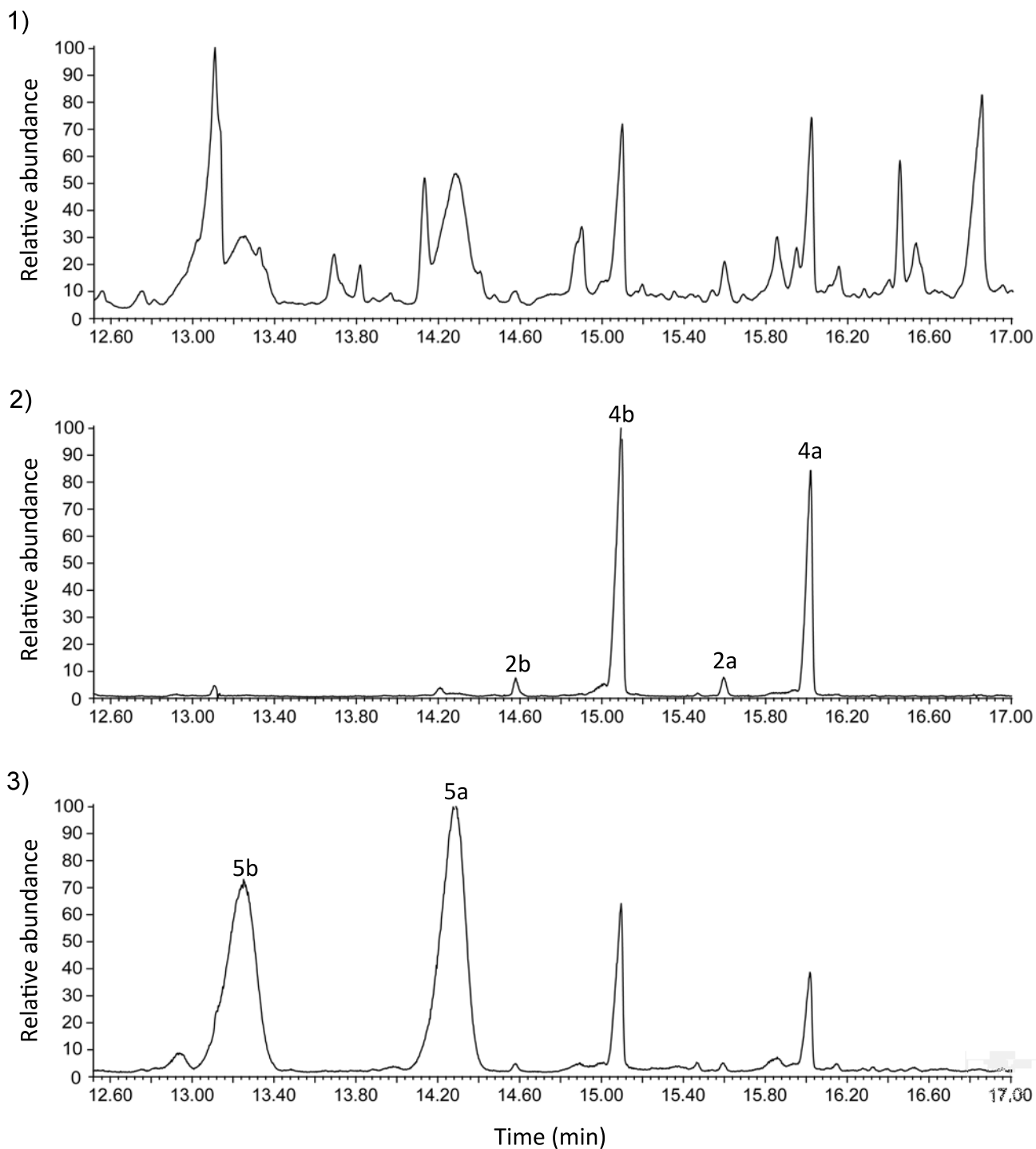
**Figure S3.** Mass spectral profiles (EI mode) of the four reduced lactones (1–4) and the ketolactone 5, where **a** refers to those derived from propionate as starter unit, and **b**, those derived from acetate. As observed previously,<sup>[8]</sup> the mass spectra of the reduced lactones are essentially identical, regardless of the stereochemistry.

Figure S4

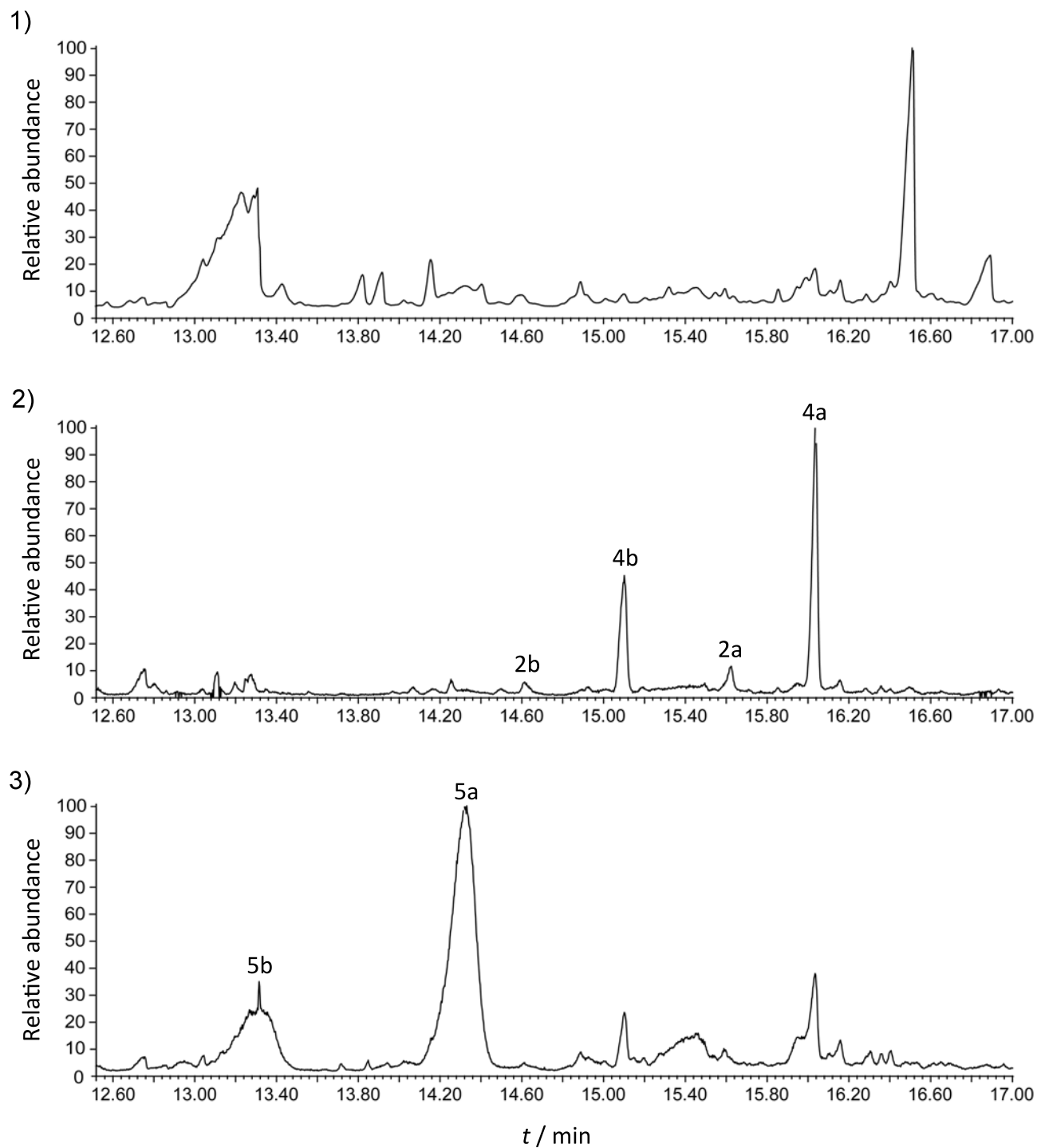


**Figure S4.** Calibration of lactone yields using standard 2. A) Calibration curve for the concentration of triketide lactone (mg L<sup>-1</sup>) vs. relative GC-MS peak area. B) Equation used to determine lactone yields in biological samples based on the approximately linear region of A).

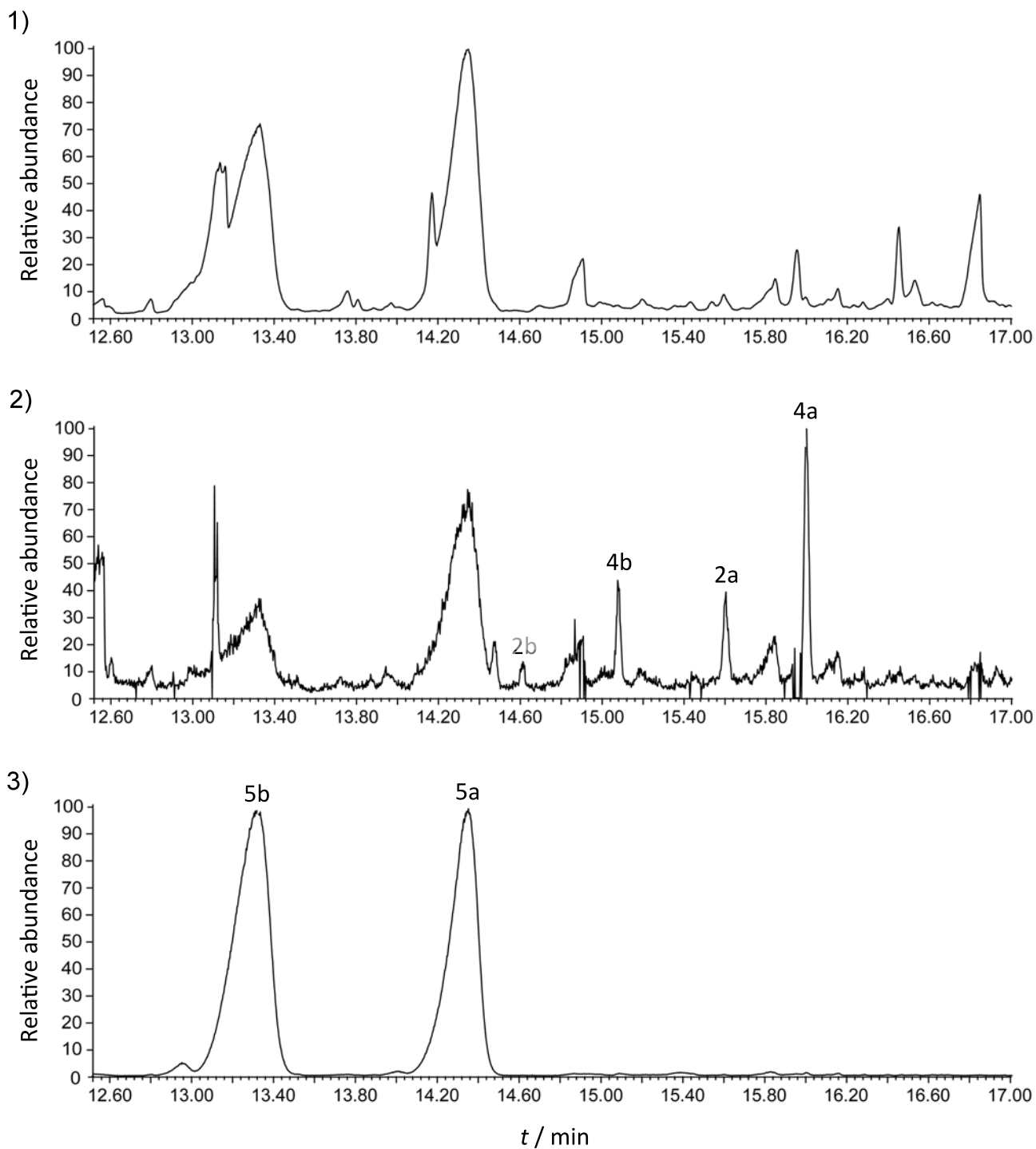




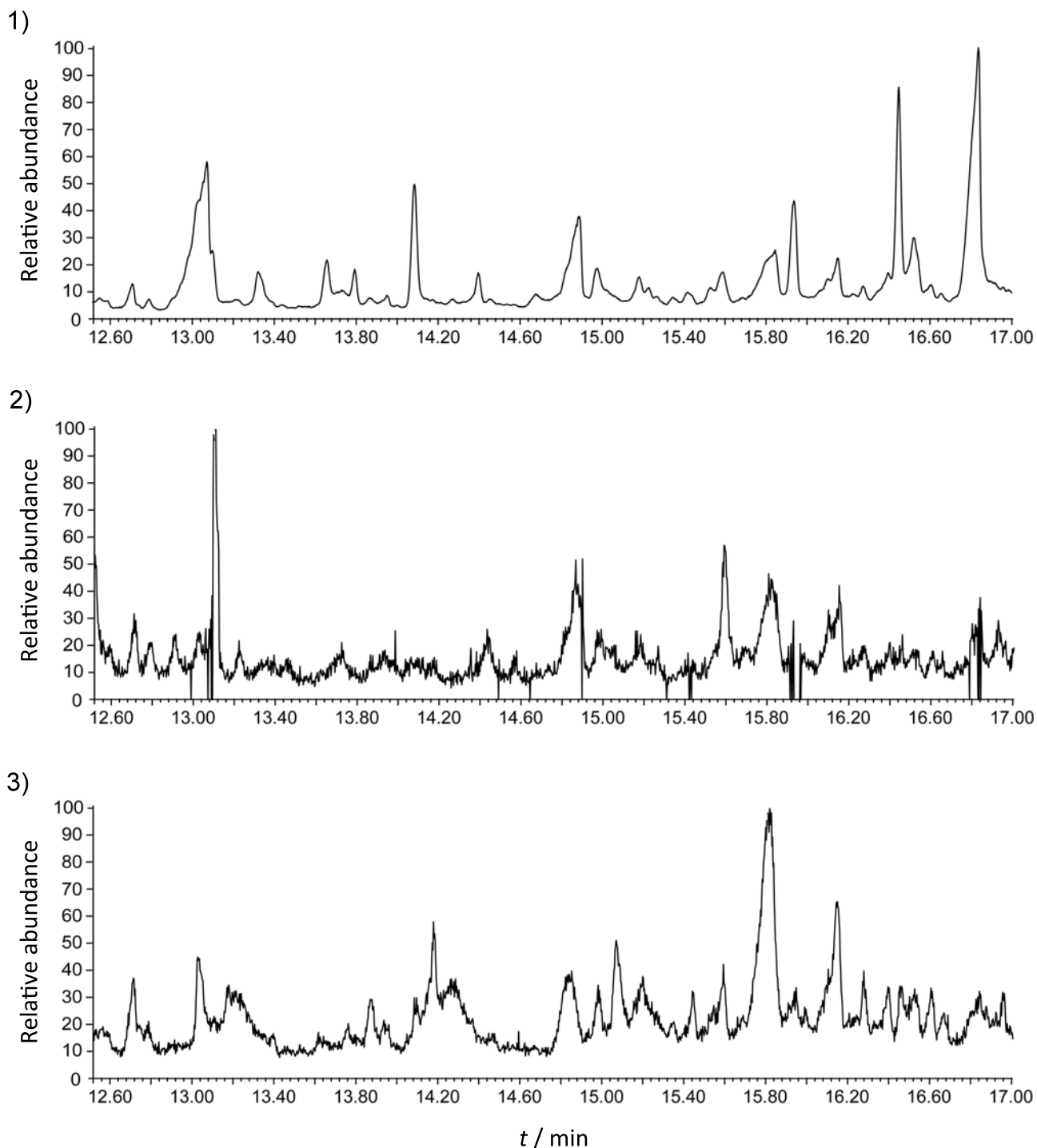
**Figure S5. A)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2, containing a hybrid DEBS 1-TE housing a reductive loop derived from amphotericin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



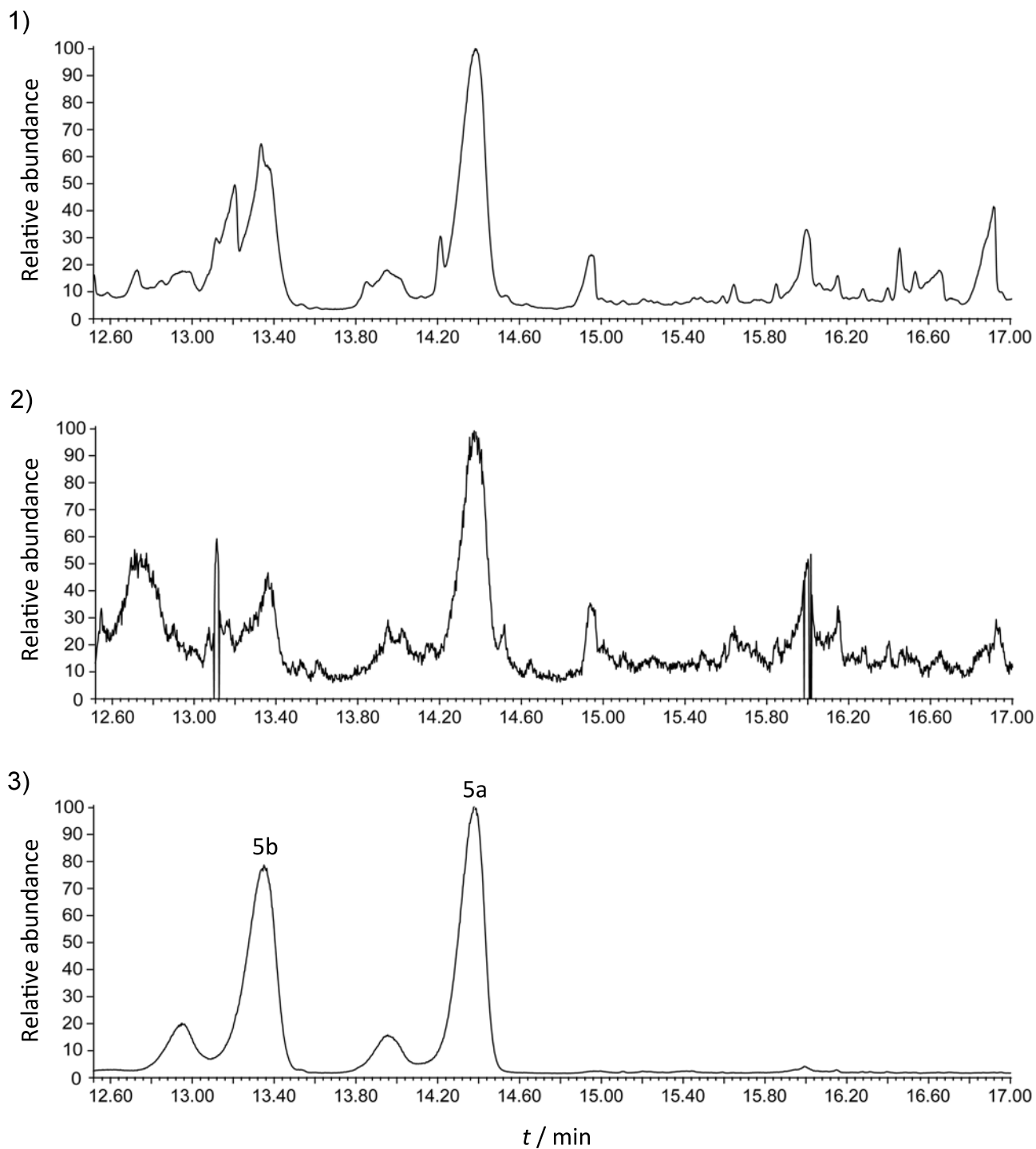
**Figure S5. B)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from amphotericin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



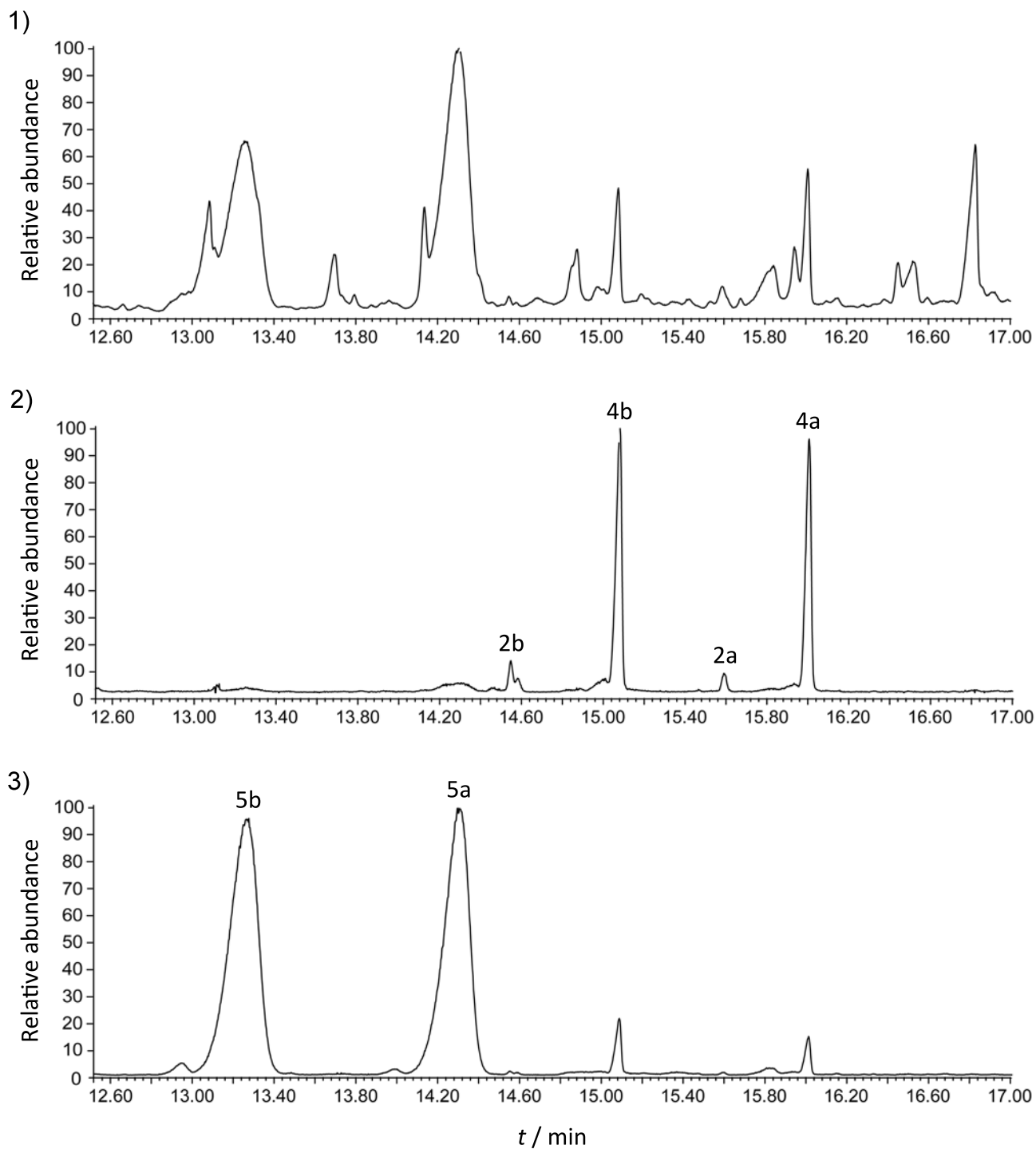
**Figure S5. C)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from concanamycin PKS module 4. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



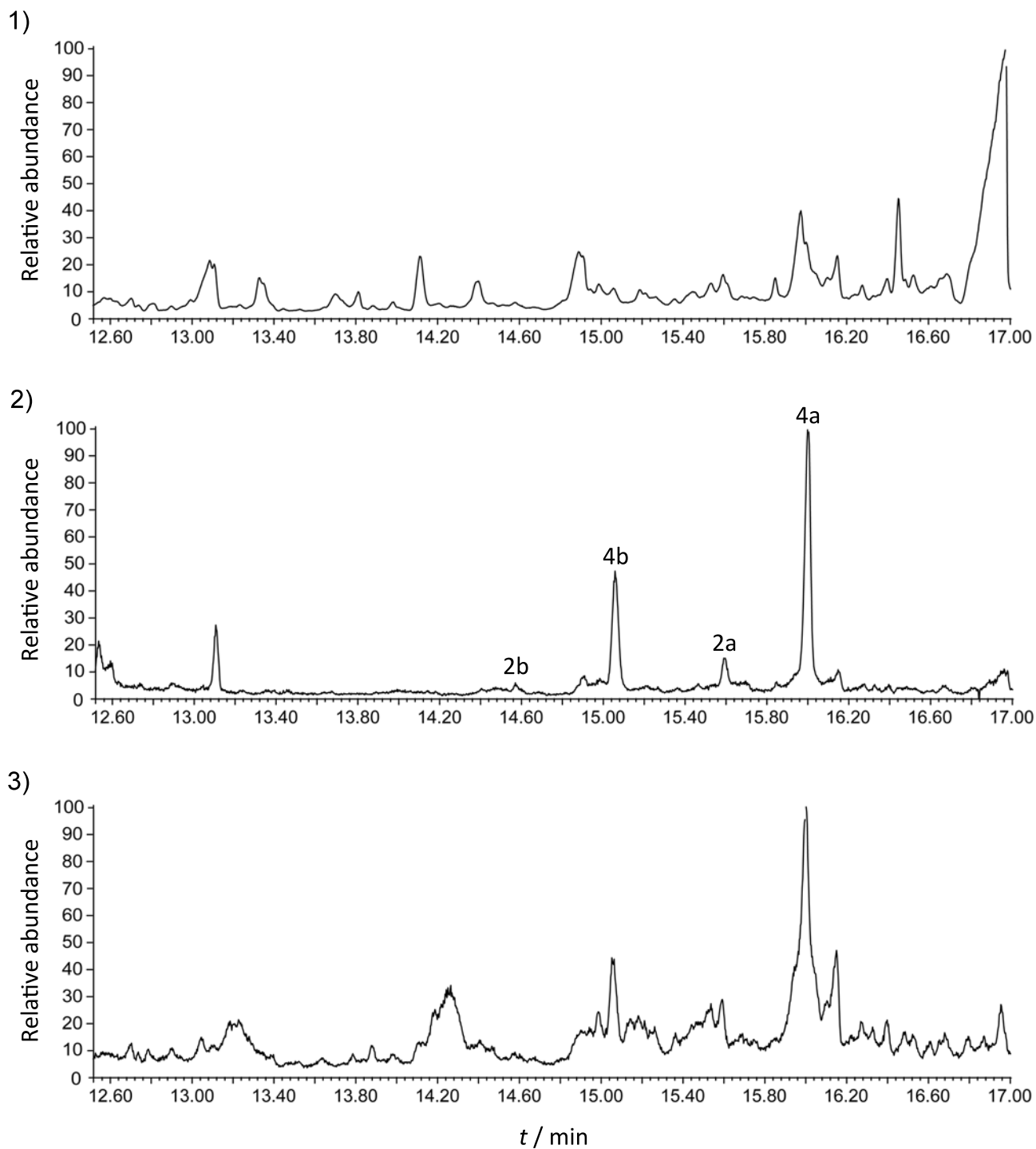
**Figure S5. D)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from elaiophyllin PKS module 4. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



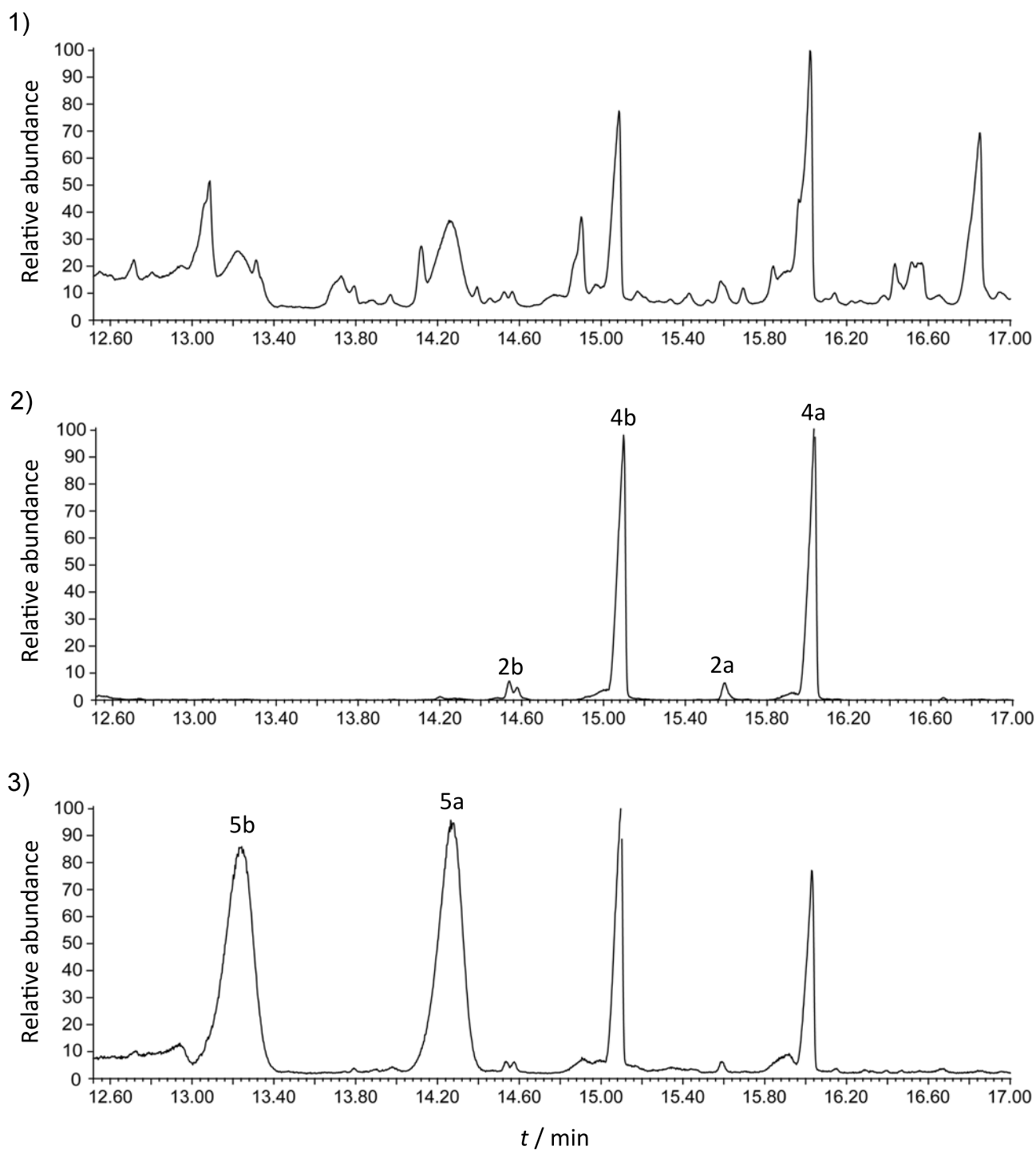
**Figure S5. E)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from oligomycin PKS module 5. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



**Figure S5. F)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from pimarin PKS module 7. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).

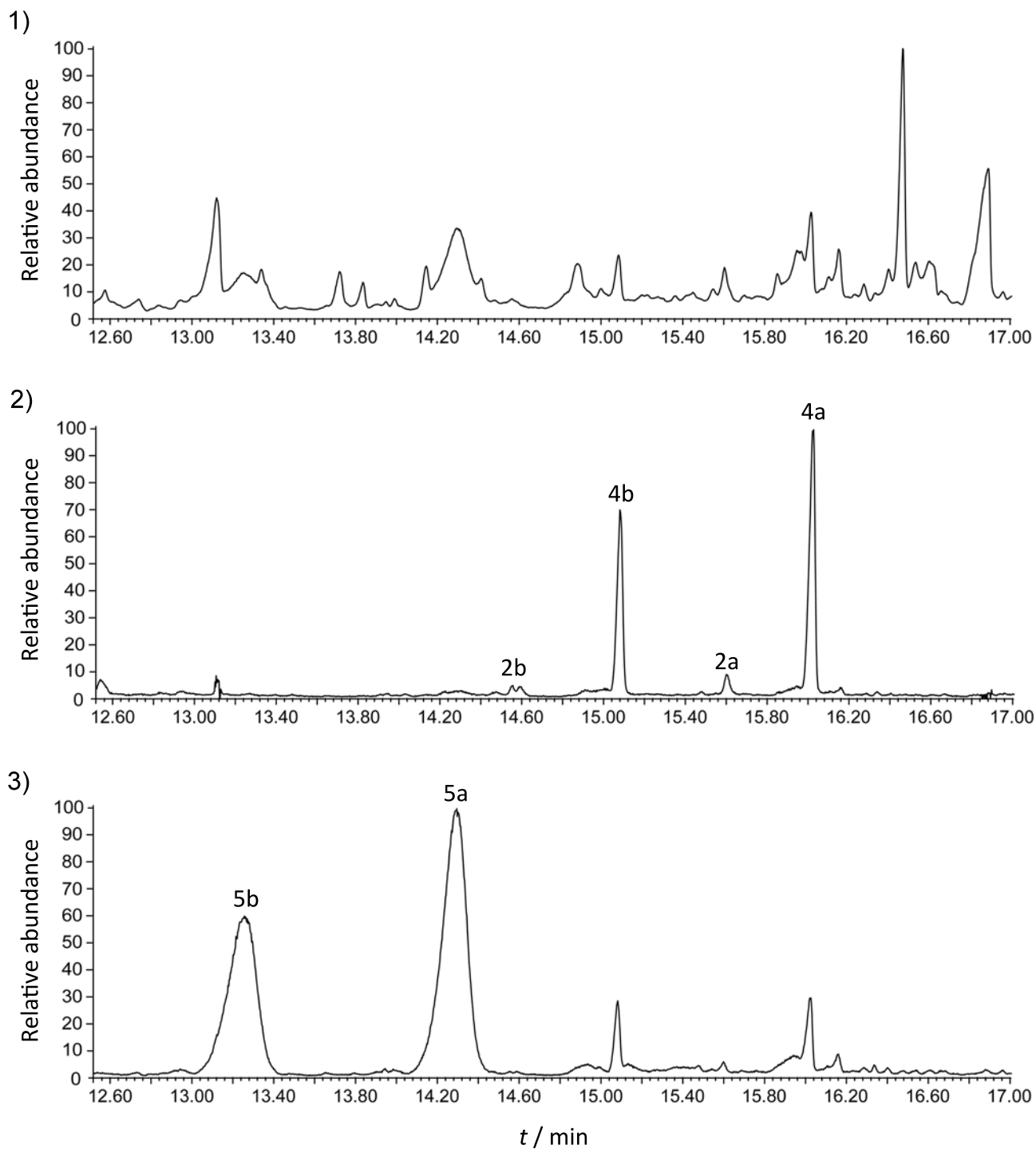


**Figure S5. G)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from concanamycin PKS module 10. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).

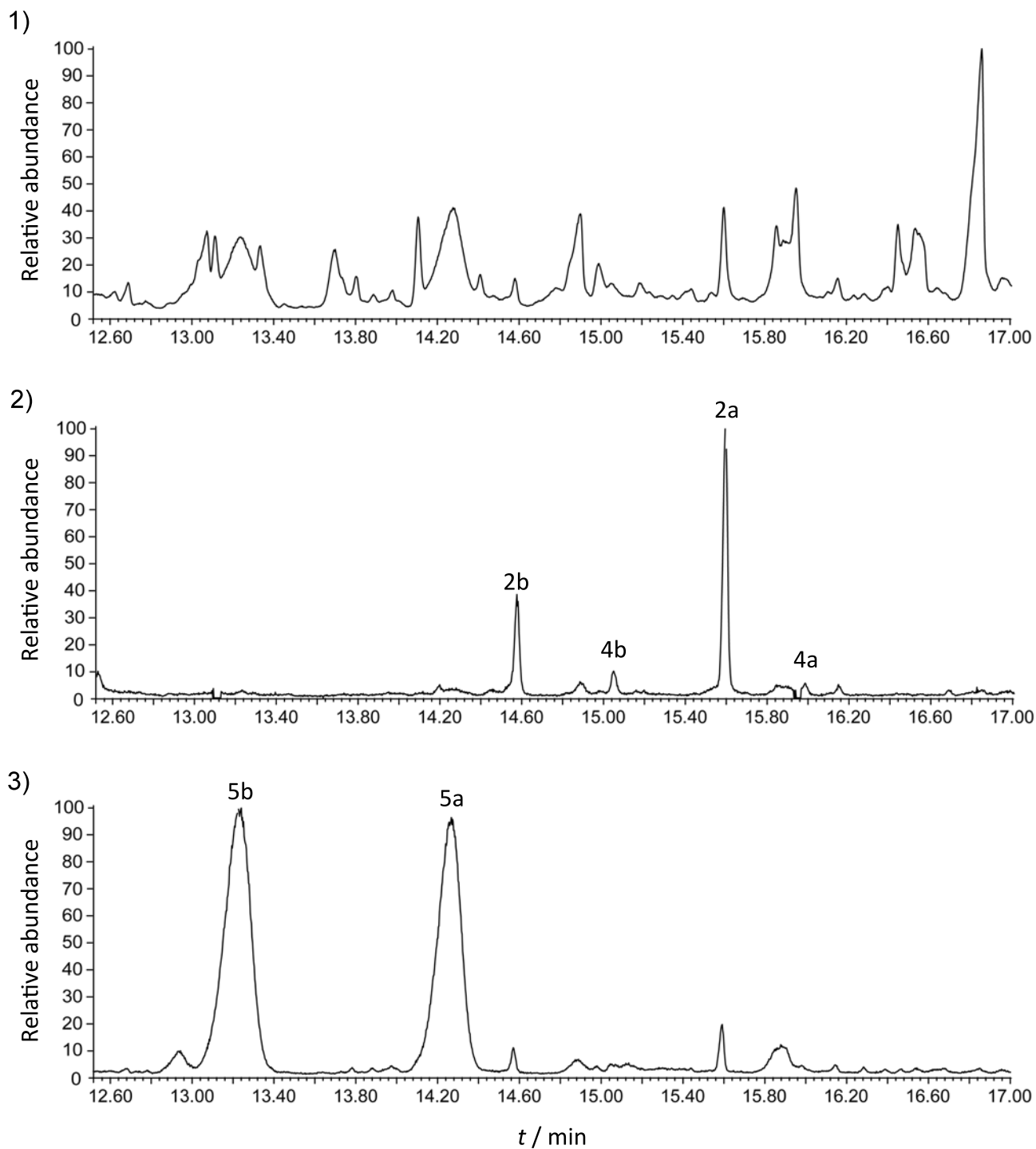


**Figure S5. H)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from amphotericin PKS module 11. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).

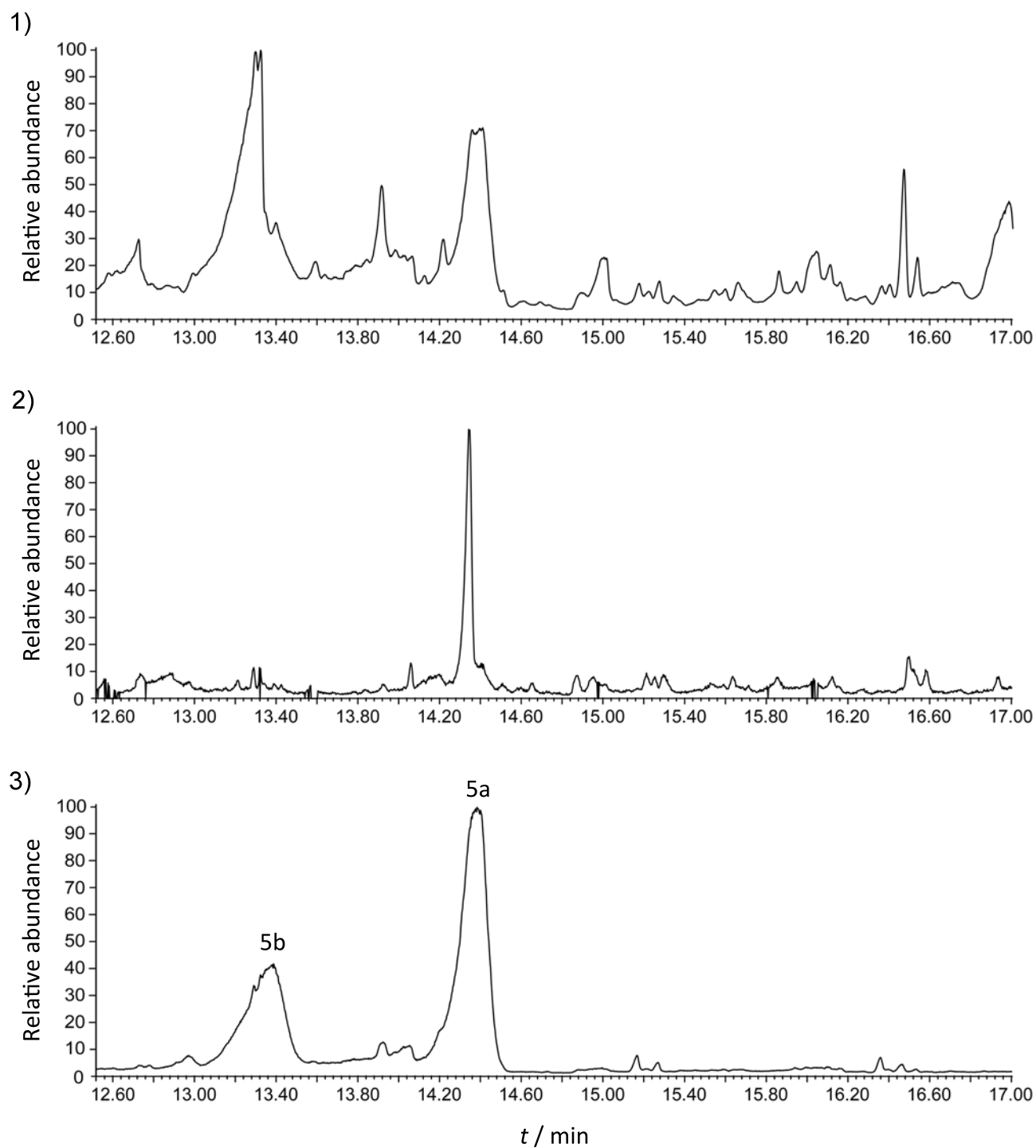




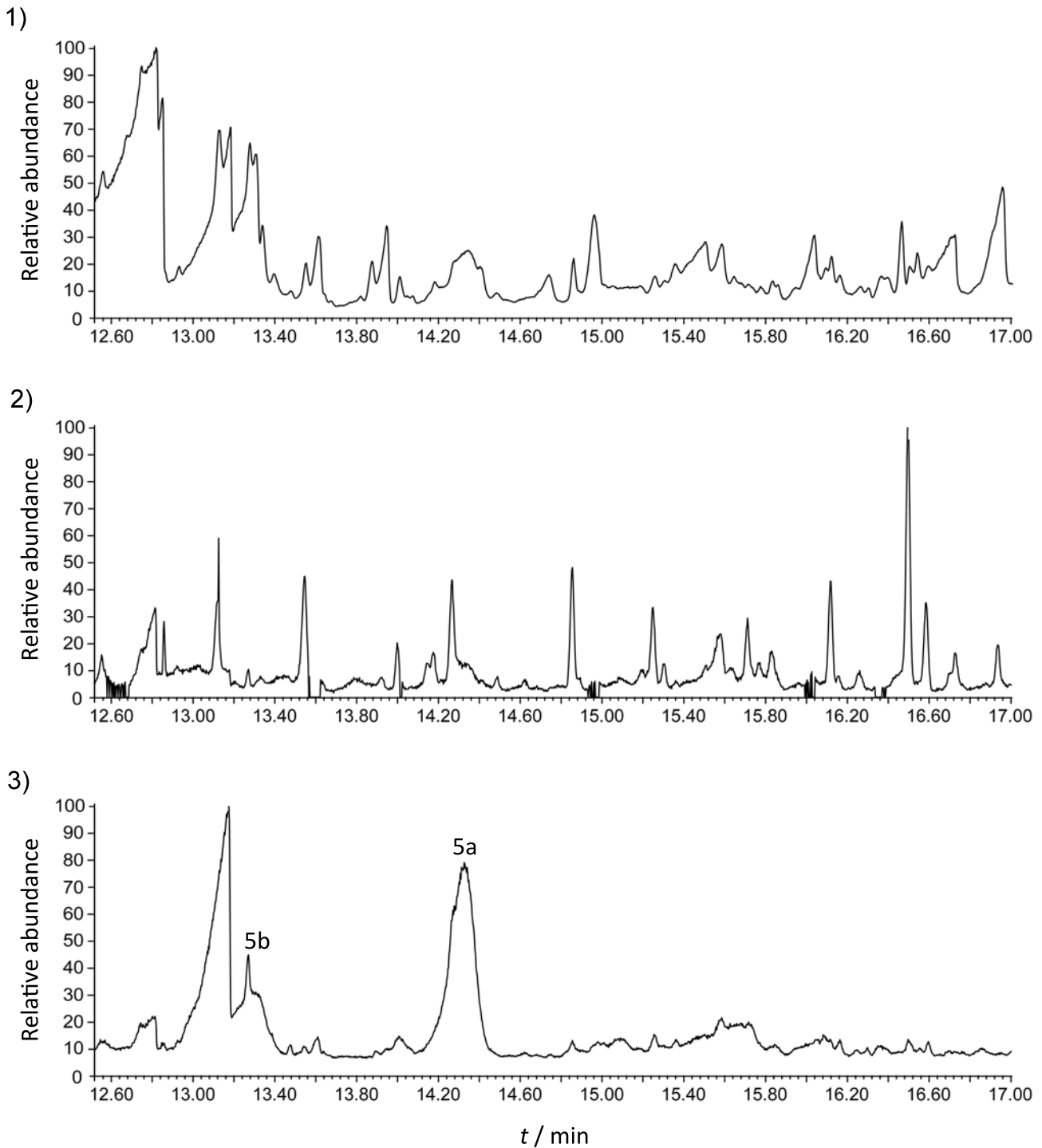
**Figure S5. I)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from candidicin PKS module 13. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



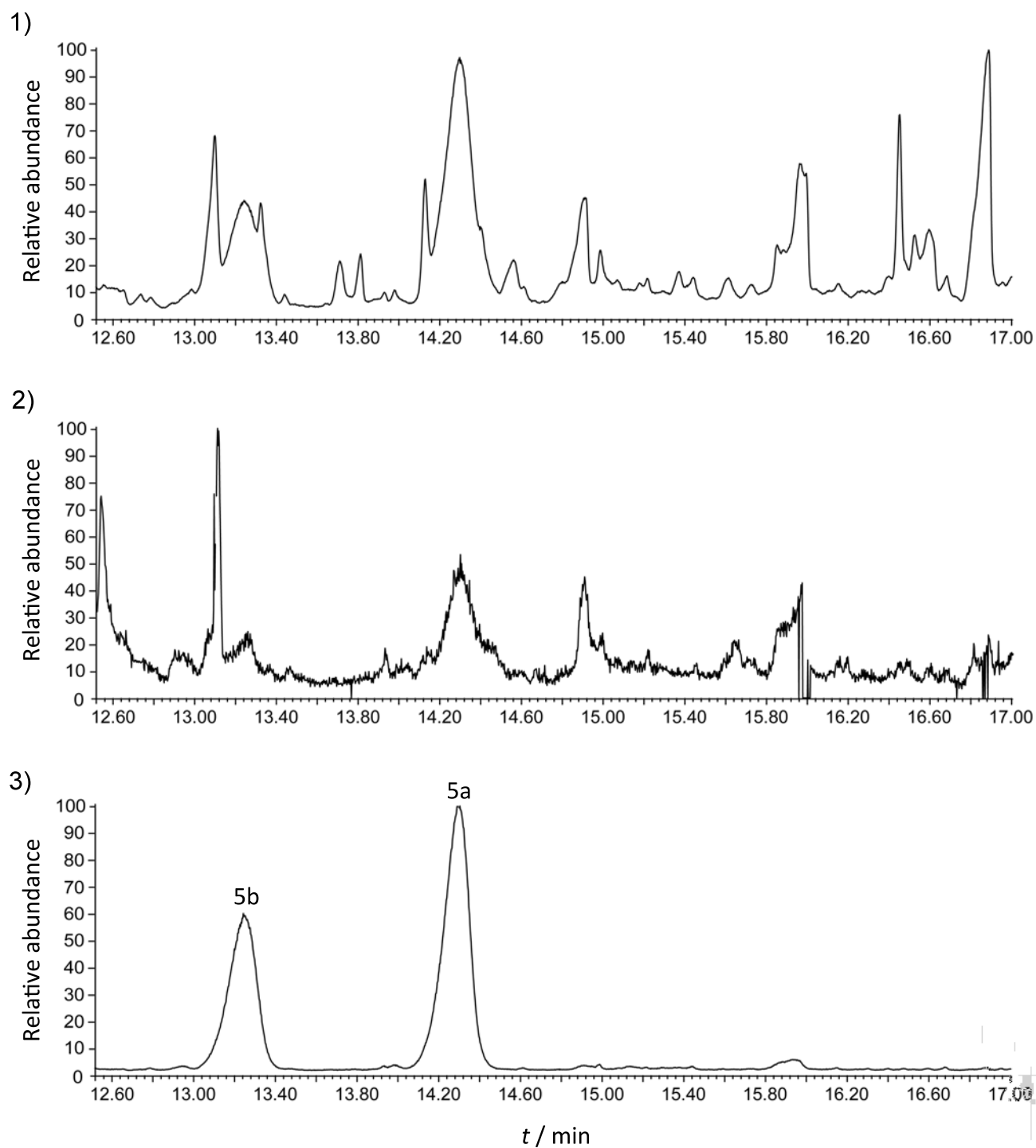
**Figure S5. J)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from the mycolactone PKS. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



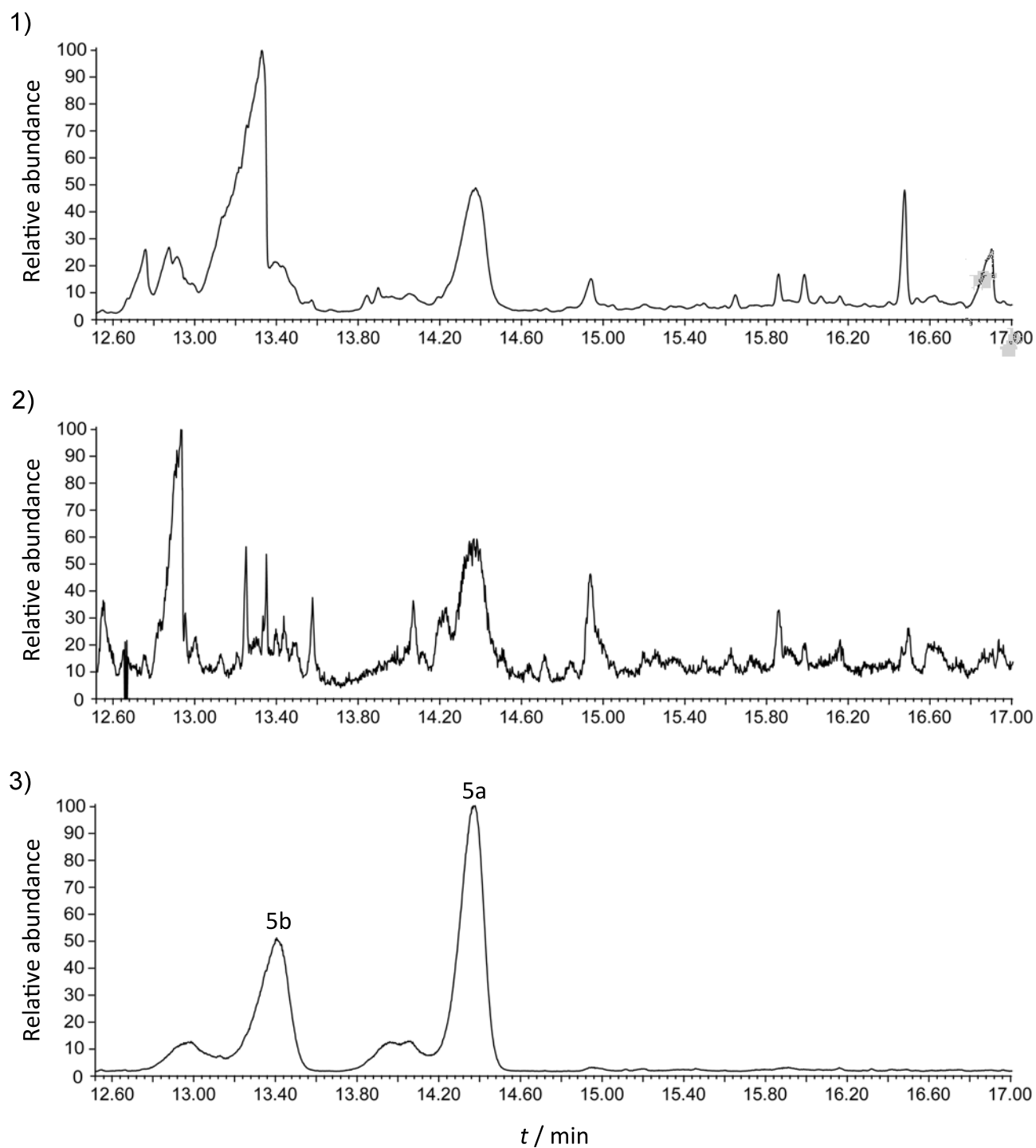
**Figure S5. K)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from erythromycin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



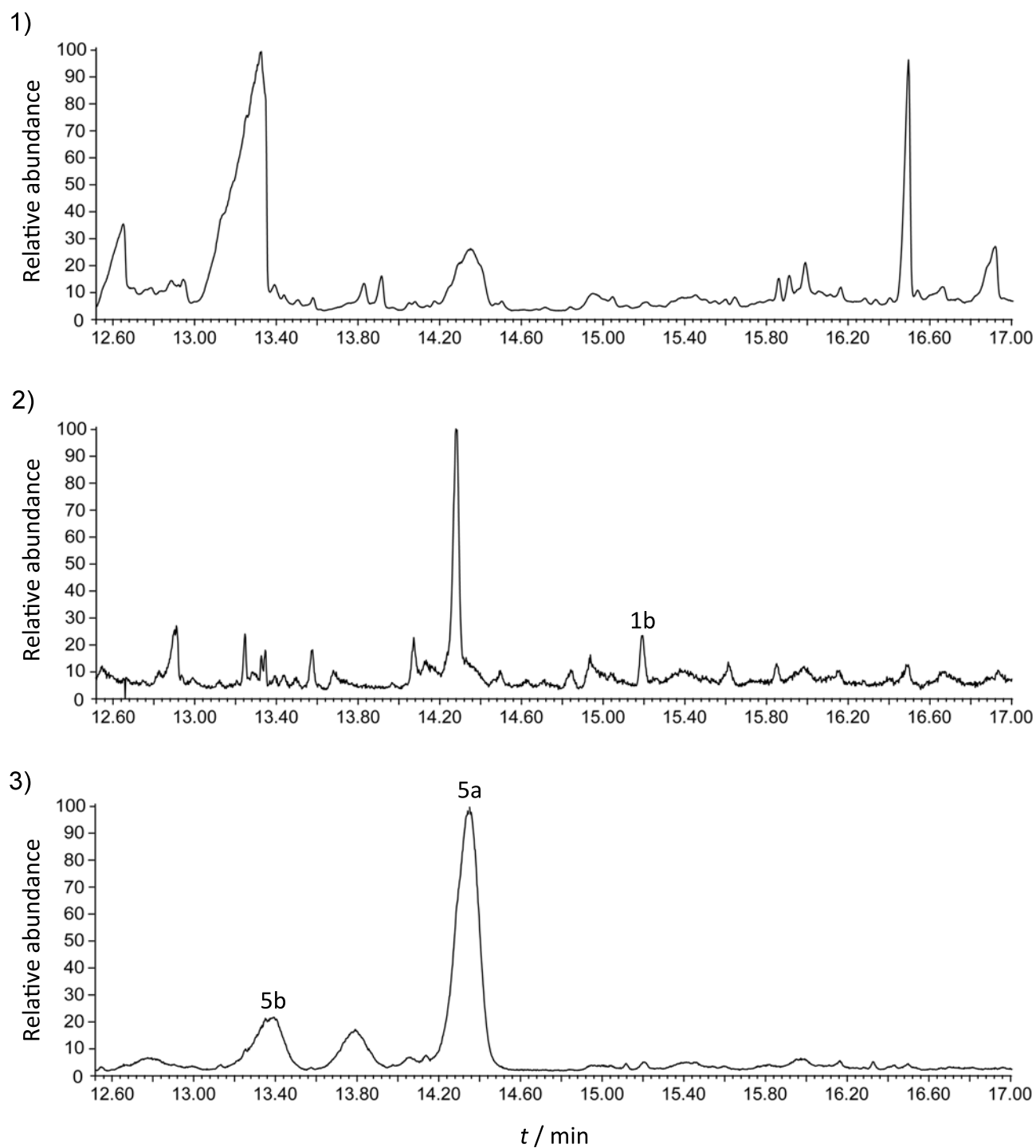
**Figure S5. L)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from erythromycin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



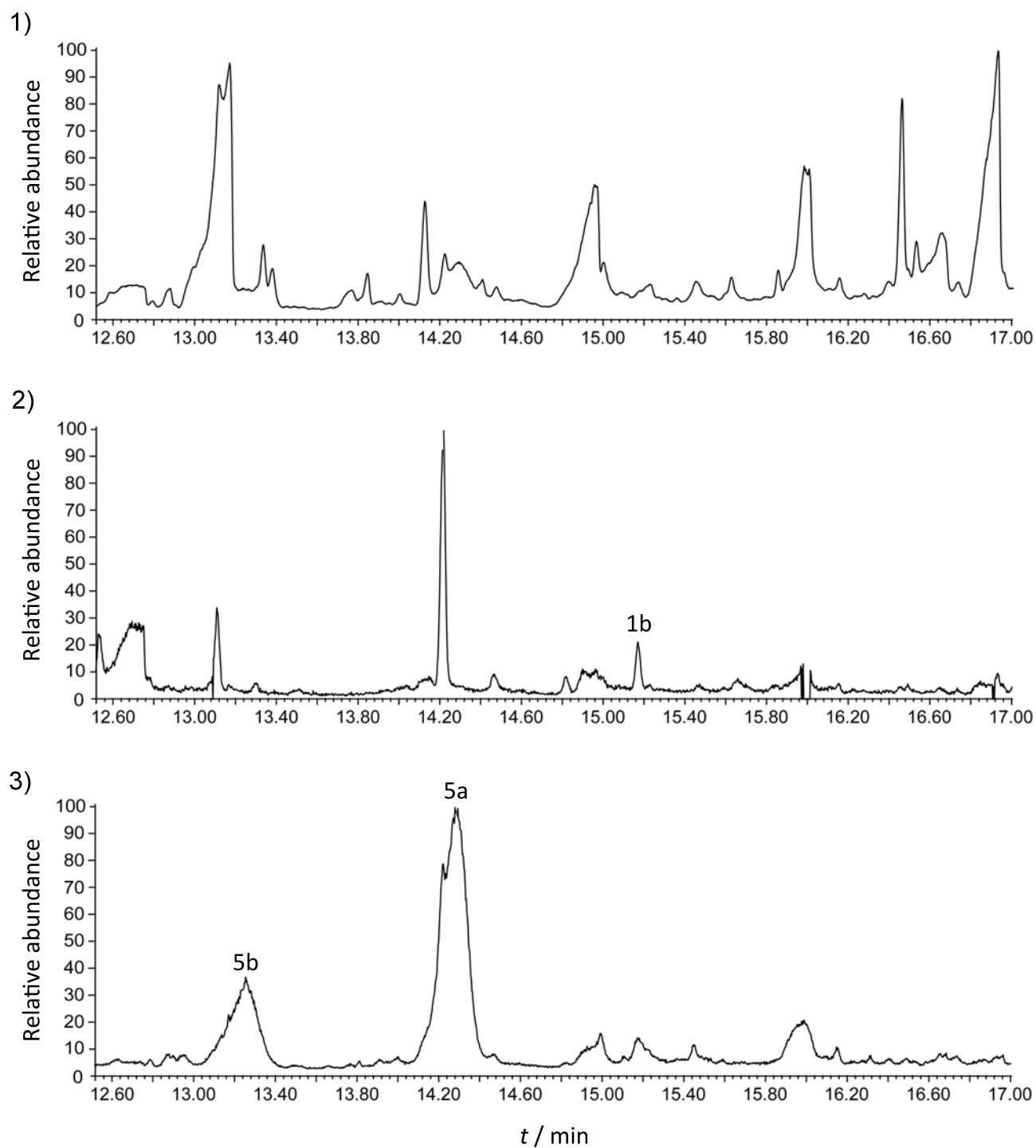
**Figure S5. M)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from pikromycin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



**Figure S5. N)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from pikromycin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).

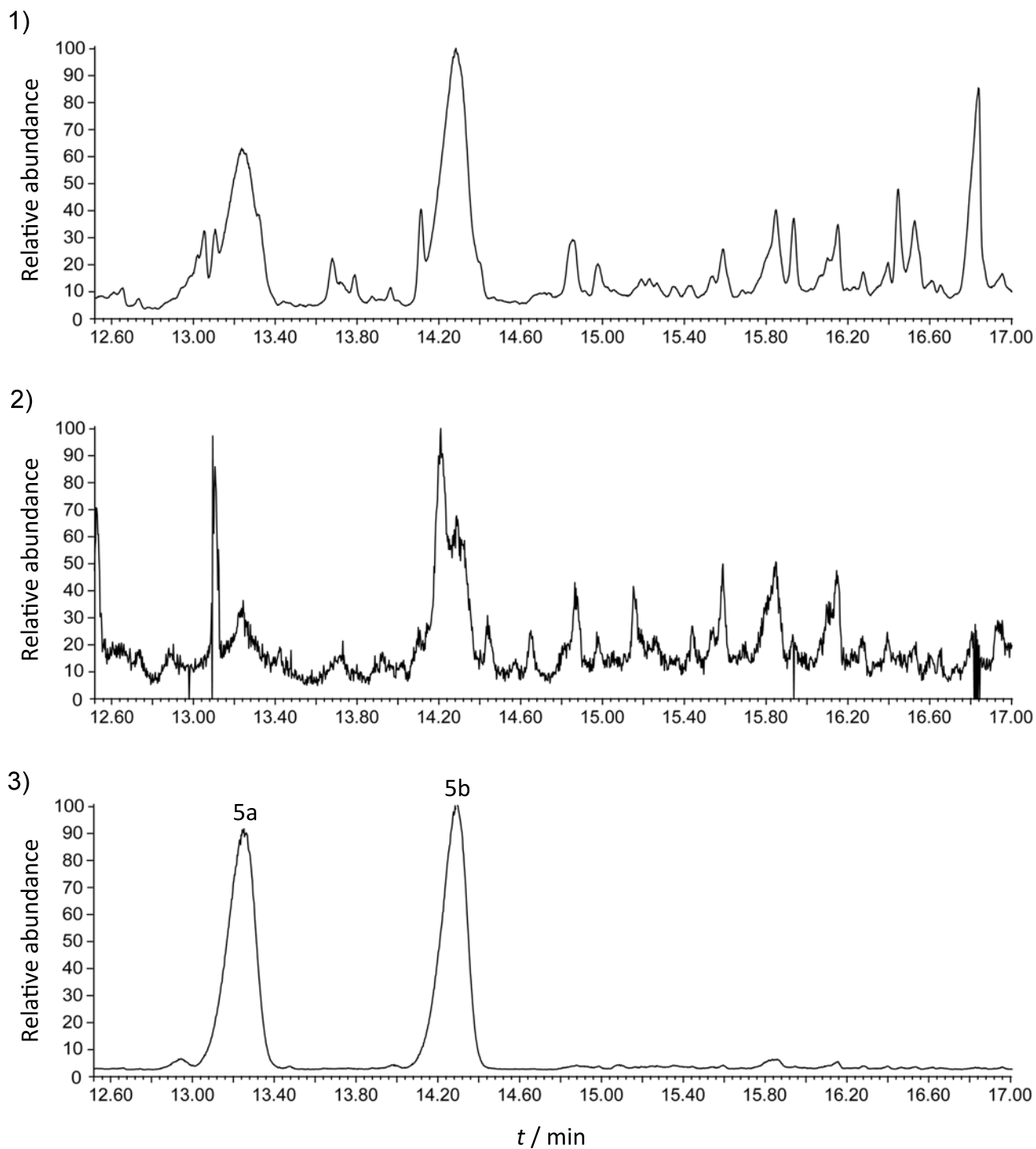


**Figure S5. O)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from lankamycin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).

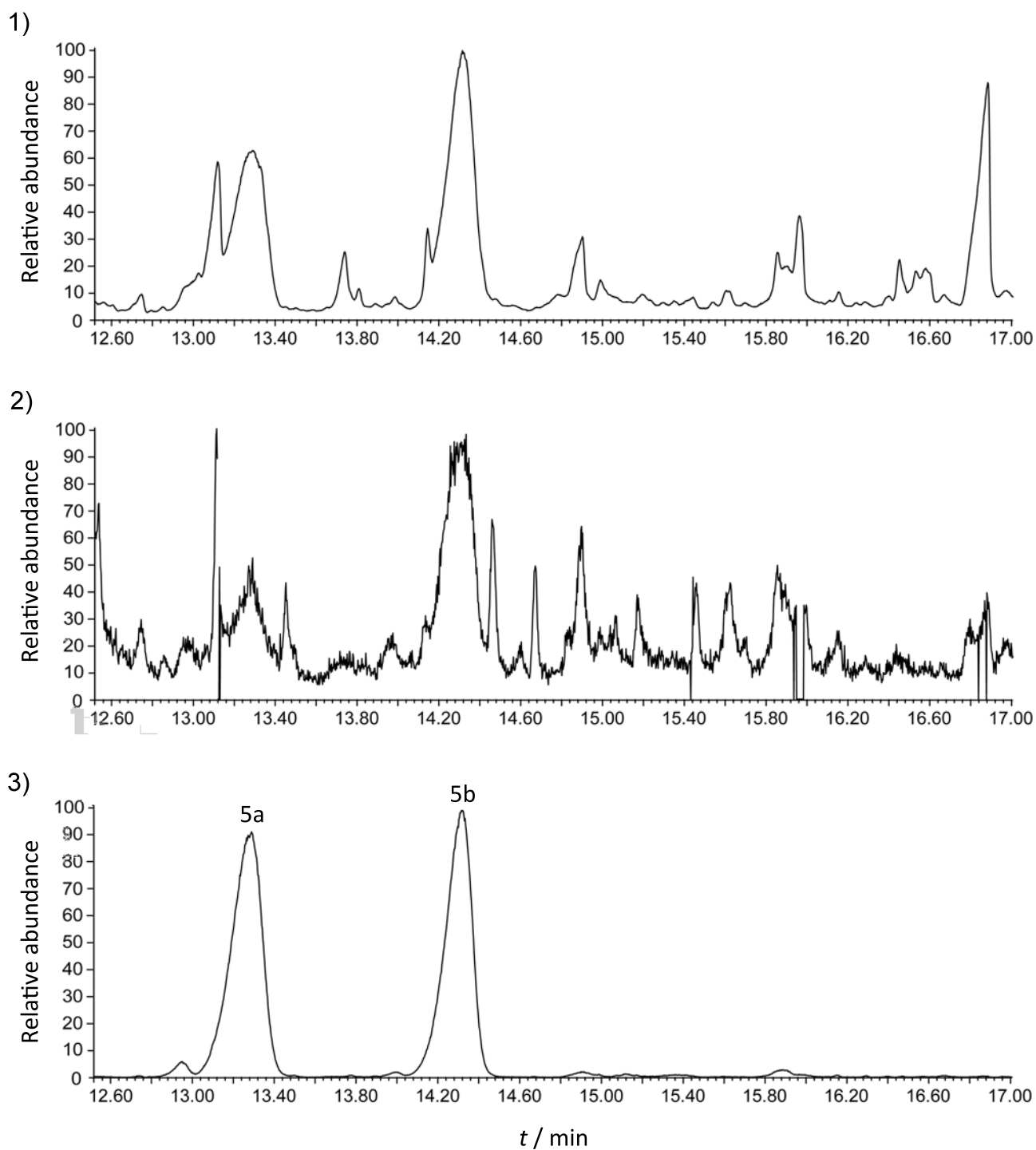


**Figure S5. P)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from lankamycin PKS module 1. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).

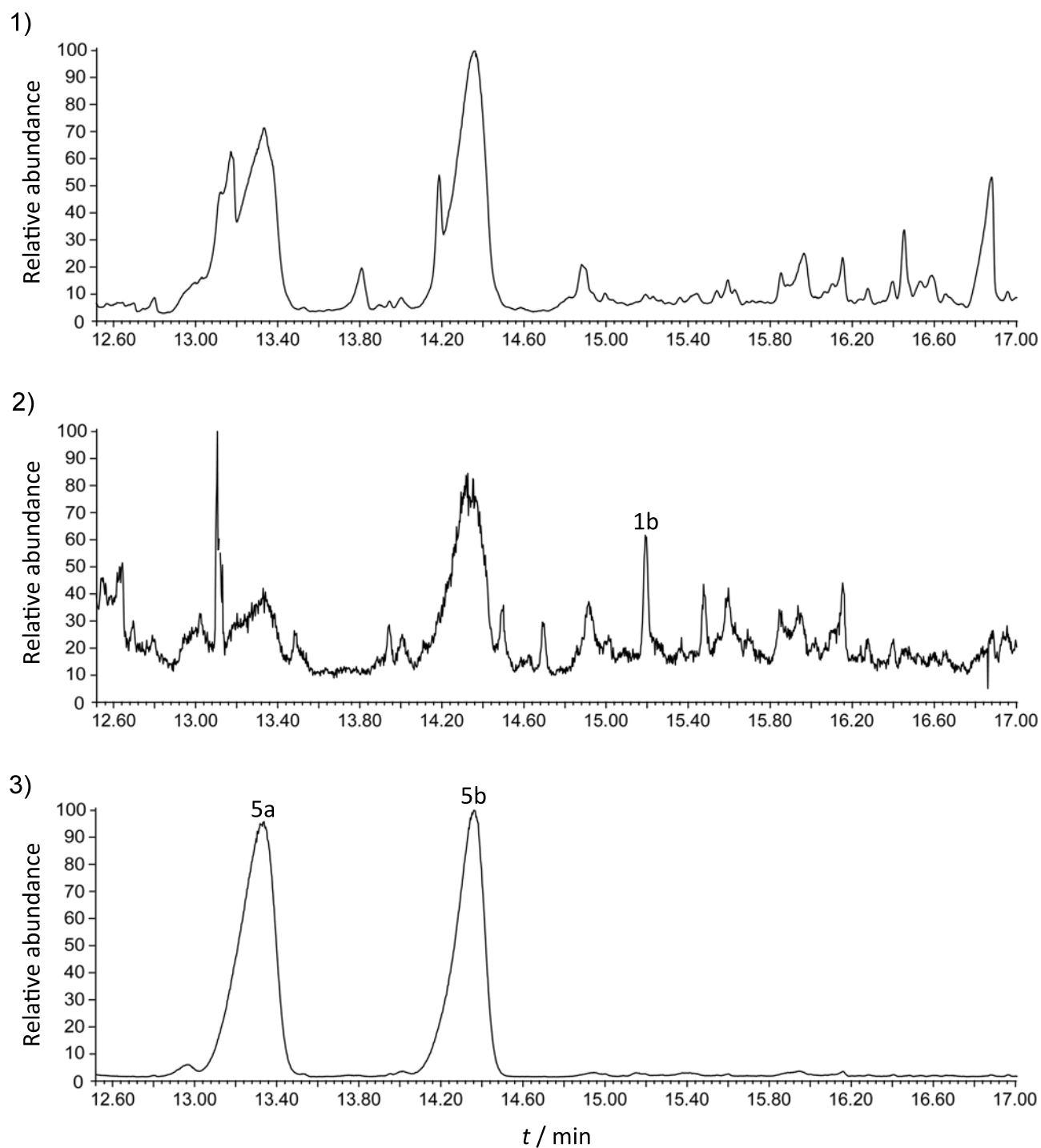




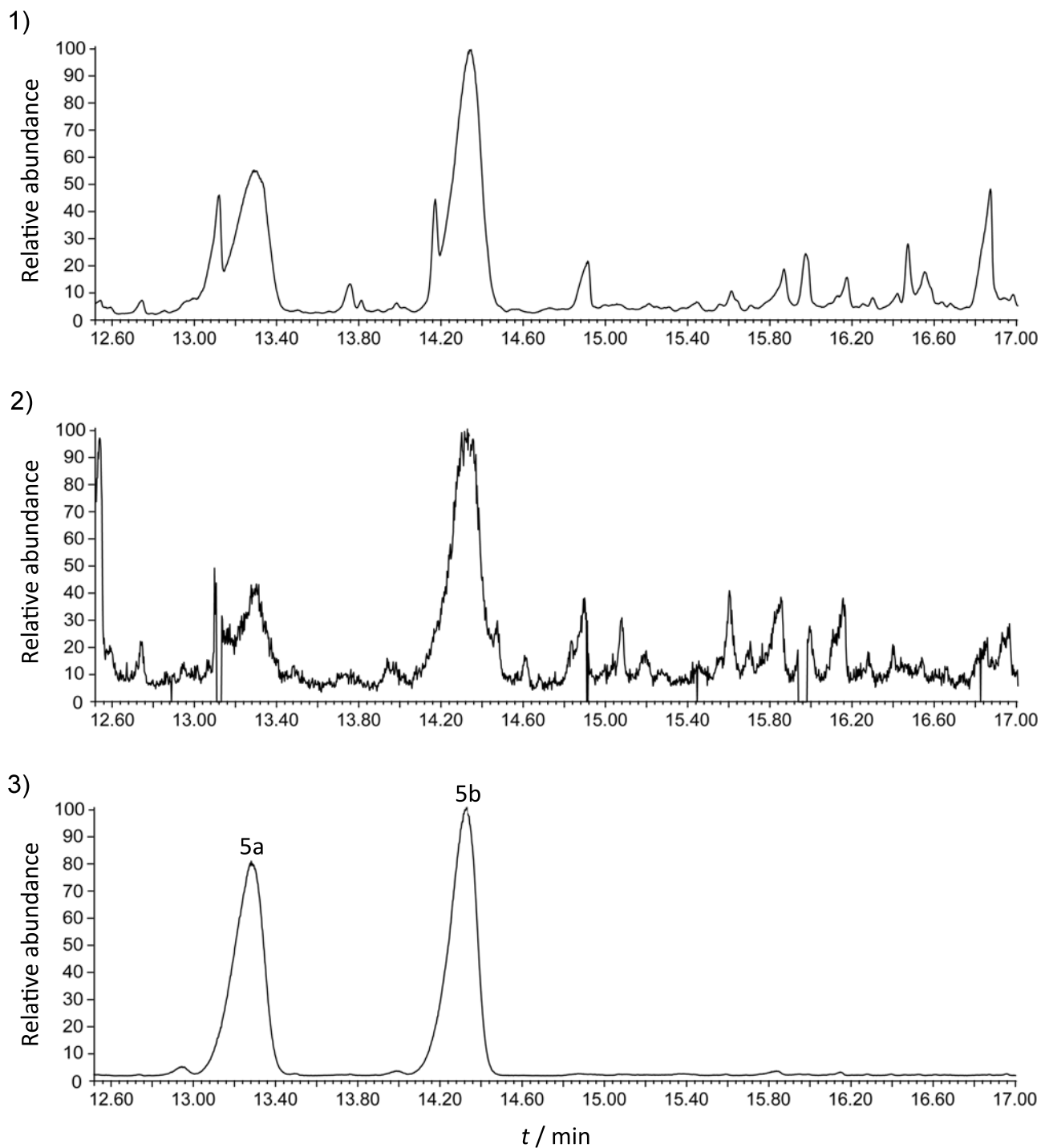
**Figure S5. Q)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from lasalocid PKS module 7. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



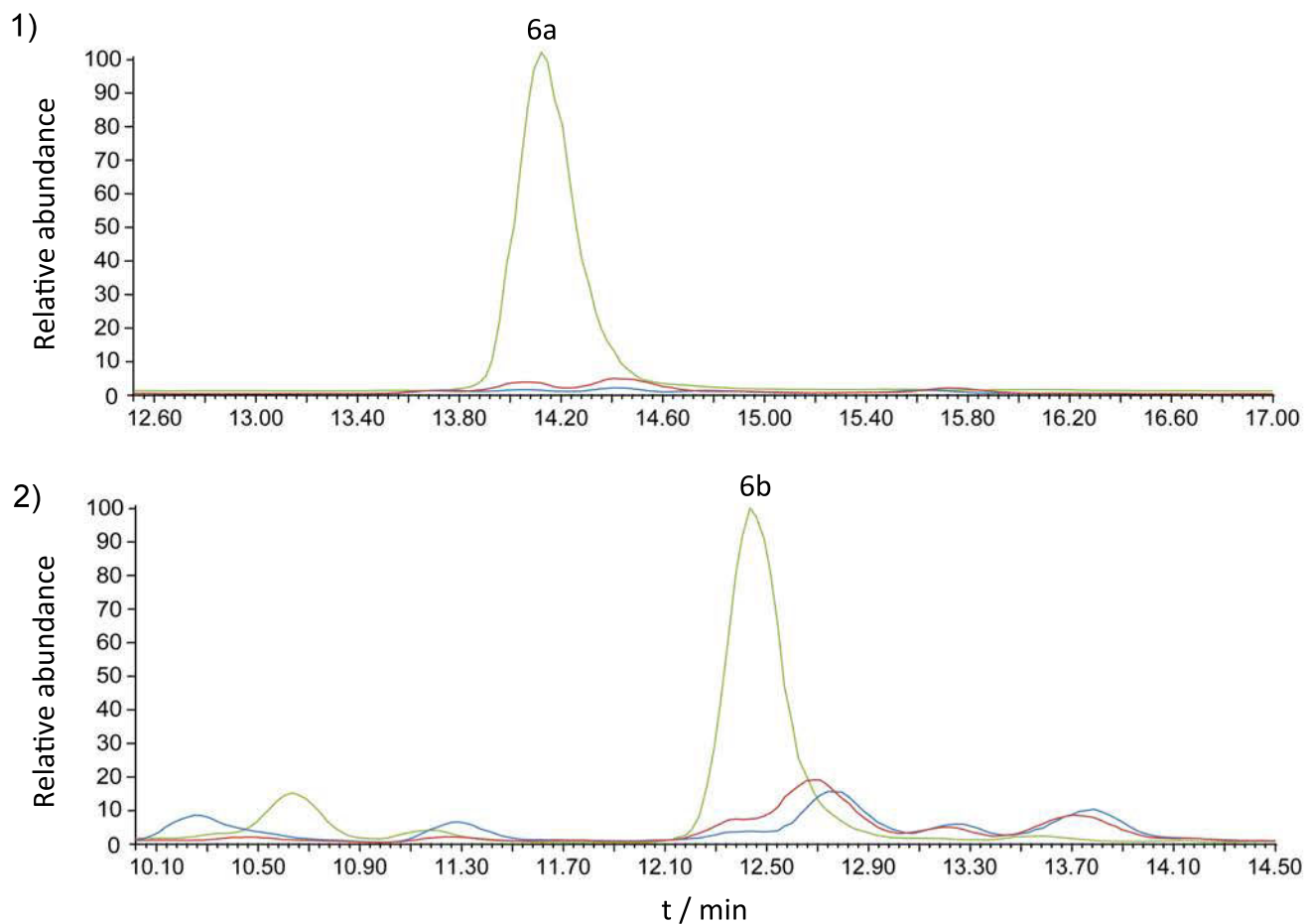
**Figure S5. R)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from ECO-02301 PKS module 19. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



**Figure S5. S)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from stambomycin PKS module 21. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



**Figure S5. T)** GC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from mycolactone PKS module 2. 1) Total Ion Chromatogram. 2) Single Ion Chromatogram (58 Da). 3) Single Ion Chromatogram (56 Da).



**Figure S5. U)** LC-MS separation and identification of triketide lactones from fermentation of *S. erythraea* JC2 containing a hybrid DEBS 1-TE housing a reductive loop derived from rapamycin PKS module 10 (green), MIs KR<sub>B</sub> (red) and MIs KR<sub>A</sub> (blue). 1) Single Ion chromatogram (187 Da). 2) Single Ion chromatogram (173 Da).

## 2. Supplementary Tables

**Table S1.** Sources of DNA used in this work

KR domain	Source of DNA
Amph KR <sub>1</sub>	<i>Streptomyces nodosus</i> strain obtained from Prof. P. Caffrey (UCD, Dublin, Ireland)
Lkm KR <sub>1</sub>	Cosmid 14F1 obtained from Prof. K. Arakawa (Hiroshima University, Japan)
Can KR <sub>13</sub>	<i>Streptomyces griseus</i> strain obtained from Prof. J. A. Gil (University of León, Spain)
DEBS KR <sub>1</sub>	Plasmid pIB023 <sup>[9]</sup>
Pik KR <sub>1</sub>	<i>Streptomyces venezuelae</i> strain obtained from Prof. Y. J. Yoon (Ewha Womans University, Seoul, Republic of Korea)
PIMS KR <sub>7</sub>	<i>Streptomyces natalensis</i> strain obtained from Prof. J. F. Aparicio (University of León, Spain)
Las KR <sub>7</sub>	<i>Streptomyces lasaliensis</i> <sup>[10]</sup>
Con KR <sub>s</sub>	<i>Streptomyces neyagawens</i> <sup>[11]</sup>
MLS KR <sub>s</sub>	BAC Mu0022B04 <sup>[5]</sup>
Eco KR <sub>19</sub>	<i>Streptomyces aizunensis</i> strain obtained from USDA, ARS culture collection (Peoria, USA)
Ela KR <sub>4</sub>	<i>Streptomyces</i> sp. DSM 4137 <sup>[12]</sup>
Sam KR <sub>21</sub>	Genomic DNA obtained from Prof. B. Aigle (University of Lorraine, France)
Olm KR <sub>5</sub>	<i>Streptomyces avermitilis</i> strain obtained from the DSMZ (Germany)

**Table S2.** Oligonucleotide primers used in this work

KR domain	Introduced restriction site	Oligonucleotide
Amph KR <sub>1</sub>	<i>AvrII</i>	5'-CCTAGGCACCGGCGCGACCCGGGTGCG-3'
	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGCAC-3'
	<i>NheI</i>	5'-GCTAGCCGGGCGGCCAGGCCGCTGGTG-3'
Lkm KR <sub>1</sub>	<i>AvrII</i>	5'-CCTAGGCACGGCGGCCCGCGCGGTGCCGCTG-3'
	<i>BglII</i>	5'-AGATCTGCCACCGCGCCGTTCCAGCGGAACGCC-3'
	<i>NheI</i>	5'-GCTAGCCCGCCGGCCCGGCCGCGGGTGCG-3'
DEBS KR <sub>1</sub>	<i>AvrII</i>	5'-CCTAGGCACCGGAGCACGCCGGG-3'
	<i>BglII</i>	5'-AGATCTGCCACCTACCGTTCCAGCGGAGCGCGTCTGG-3'
	<i>NheI</i>	5'-GCTAGCGCGCCACCCGCGGTTCC-3'
Pik KR <sub>1</sub>	<i>AvrII</i>	5'-CCTAGGCACGGCCTCCCGCCCCGGTCTGCCC-3'
	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGGCCGAG-3'
	<i>NheI</i>	5'-GCTAGCCCTGGGCGGAGCTCCCGCCCTGTCCGGACG-3'
Amph KR <sub>11</sub>	<i>BglII</i>	5'-AGATCTGCCACCTATGCCTTCCAGCACGAGAACC-3'
	<i>NheI</i>	5'-GCTAGCGCGGCGGCCGAACCTCGCCCCGGG-3'
Can KR <sub>13</sub>	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGCGTGAGCACCTG-3'
	<i>NheI</i>	5'-GCTAGCCGGTGGATGAACTTCCGGCCTCGCCGG-3'
PIMS KR <sub>7</sub>	<i>BglII</i>	5'-AGATCTGCCACGTACGCCTTCCAGCGGAGCACCTGTGGG-3'
	<i>NheI</i>	5'-GCTAGCCGGGTGGCGAACTCGCTGCCGG-3'
Las KR <sub>7</sub>	<i>BglII</i>	5'-AGATCTGCCGACGTACGCCTTCCAGCGTCAGCG-3'
	<i>NheI</i>	5'-GCTAGCCGGTTCGAGGAGTCCGAC-3'
Con KR <sub>10</sub>	<i>BglII</i>	5'-AGATCTCCACCTACCCCTTCCGAACACCGG-3'
	<i>NheI</i>	5'-GCTAGCCGCTCGCGCAGCACGTCGGAGGCCG-3'
Con KR <sub>4</sub>	<i>BglII</i>	5'-AGATCTGCCACGTACCCCTTCCAGCGGCGC-3'
	<i>NheI</i>	5'-GCTAGCCGGTTCGCGCAGCGAGGAGCGGTGTC-3'
MIs KR <sub>A</sub>	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGCGGC-3'
	<i>NheI</i>	5'-GCTAGCTTGCTCGTCAGGACGTCGC-3'
MIs KR <sub>B</sub>	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGCATCAGCGTACTGGCTGCTG-3'
	<i>NheI</i>	5'-GCTAGCCGGGCGAGGATCGGTTGCGGCAGAGGC-3'
Eco KR <sub>19</sub>	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGCAGGAG-3'
	<i>NheI</i>	5'-GCTAGCTGCTGGGCCAGCGCGGC-3'
Ela KR <sub>4</sub>	<i>BglII</i>	5'-AGATCTTGCCGACGTACGCCTTCCAGCGGCAGCGCTACTGGC-3'
	<i>NheI</i>	5'-GCTAGCCGGTTCGCGCAGCGAGGAGG-3'
Sam KR <sub>21</sub>	<i>BglII</i>	5'-AGATCTGCCACCTACGCCTTCCAGCGTTCGC-3'
	<i>NheI</i>	5'-GCTAGCCGCTGGGCCGGCGAGAGGC-3'
Olm KR <sub>5</sub>	<i>BglII</i>	5'-AGATCTGCCCGCCTACCCCTTCCAGCGCACCCGCTACTGG-3'
	<i>NheI</i>	5'-GCTAGCCGTTGGGCGAGCGGGGAGGCGGCG-3'
MIs KR <sub>B</sub> A5459V	F	5'-TAGTGCACACCGCCGTGGTATTGGCCGACAC-3'
	R	5'-GTGTGCGCCAATACCACGGCGGTGTGCACTA-3'
MIs KR <sub>B</sub> Q5515L	F	5'-TAGGCAGTCCCGGTCTGGGTAACCTACGCGGC-3'
	R	5'-GCCGCGTAGTTACCCAGACCCGGACTGCCTA-3'
MIs KR <sub>B</sub> H5550R	F	5'-GCTACTGGCAGACTCGCACCGGTCTCACCGC-3'
	R	5'-GCGGTGAGACCGGTGCGAGTCTGCCAGTAGC-3'
MIs KR <sub>B</sub> <i>BglII</i> site removal	F	5'-CTAACCGATGTAGACCTAGCCCGCATGAC-3'
	R	5'-GTCATGCGGGCTAGGTCTACATCGGTTAG-3'
MIs KR <sub>B</sub> <i>NheI</i> site removal	F	5'-CGGATTTGTGCGAGCTGGCGCTGCATGTCGCT-3'
	R	5'-AGCGACATGCAGCGCCAGCTCGACAAATCCG-3'

### 3. Supplementary References

- [1] Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11124–11129.
- [2] V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi, C. Khosla, *Protein Sci.* **2007**, *16*, 2093–2107.
- [3] A. T. Keatinge-Clay, R. M. Stroud, *Structure* **2006**, *14*, 737–748.
- [4] J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2013**, *8*, 1263–1270.
- [5] T. P. Stinear, A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay, S. T. Cole, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1345–1349.
- [6] A. T. Keatinge-Clay, *Chem. Biol.* **2007**, *14*, 898–908.
- [7] P. Caffrey, *ChemBioChem* **2003**, *4*, 654–657.
- [8] K. J. Weissman, G. C. Kearney, P. F. Leadlay, J. Staunton, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2103–2108.
- [9] R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, *Chem. Biol.* **2003**, *10*, 723–731.
- [10] L. Smith, H. Hong, J. B. Spencer, P. F. Leadlay, *ChemBioChem* **2008**, *9*, 2967–2975.
- [11] S. F. Haydock, A. N. Appleyard, T. Mironenko, J. Lester, N. Scott, P. F. Leadlay, *Microbiology* **2005**, *151*, 3161–3169.
- [12] S. F. Haydock, T. Mironenko, H. I. Ghoorahoo, P. F. Leadlay, *J. Biotechnol.* **2004**, *113*, 55–68.



## Partie 3 : Essai de résolution de la structure d'une KR épimérisante

### I. Clonage et purification des KR

Les séquences codant la KR<sub>1</sub> de la PKS lankamycine (Lank KR<sub>1</sub>), la KR<sub>1</sub> de la PKS pikromycine (Pik KR<sub>1</sub>) et la KR<sub>6</sub> de la PKS virginiamycine (Vir KR<sub>6</sub>) ont été amplifiées à partir du cosmide 14F1, de l'ADN génomique de *S. venezuelae* et de l'ADN génomique de *S. virginiae*, respectivement. Elles ont été clonées dans le vecteur pBG-106. Les vecteurs, dont la KR<sub>A</sub> clonée dans le pVB (Pr Leadlay, Université de Cambridge, UK), ont ensuite été utilisés pour transformer des cellules BL21(DE3) d'*E. coli*. La production des KR recombinantes a été réalisée en milieu ZYM 5052. Les cellules ont ensuite été lysées et le surnageant a été purifié en plusieurs étapes. La première étape a consisté en une colonne de Nickel-Sépharose. L'échantillon protéique obtenu a ensuite été dialysé en présence de protéase PreScission ou TEV pour la KR<sub>A</sub> afin de cliver le tag histidine, puis déposé sur une nouvelle colonne de Nickel-Sépharose. La purification s'est terminée par une colonne de filtration sur gel. La pureté des protéines a été vérifiée par gel SDS-PAGE 12,5 % (Figure 57). Les protéines ont été congelées dans l'azote liquide puis stockées à -80 °C. Le rendement global était de 22 mg de Lank KR<sub>1</sub>/L de culture, 7 mg de Pik KR<sub>1</sub>/L de culture, 14 mg de Vir KR<sub>6</sub> /L de culture pour la et 8 mg de Mls KR<sub>A</sub> /L de culture.

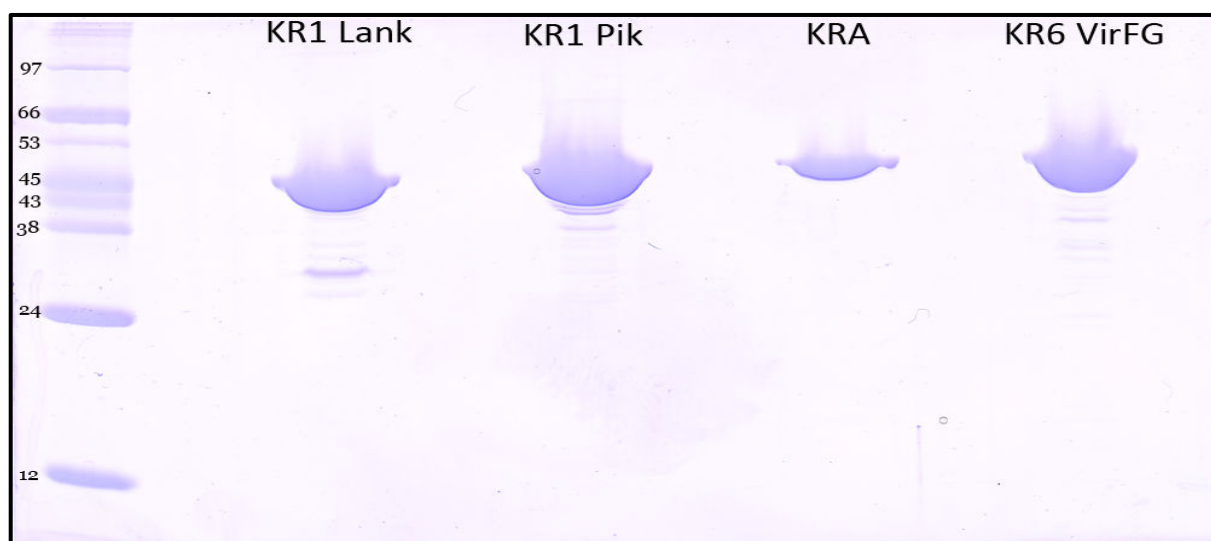
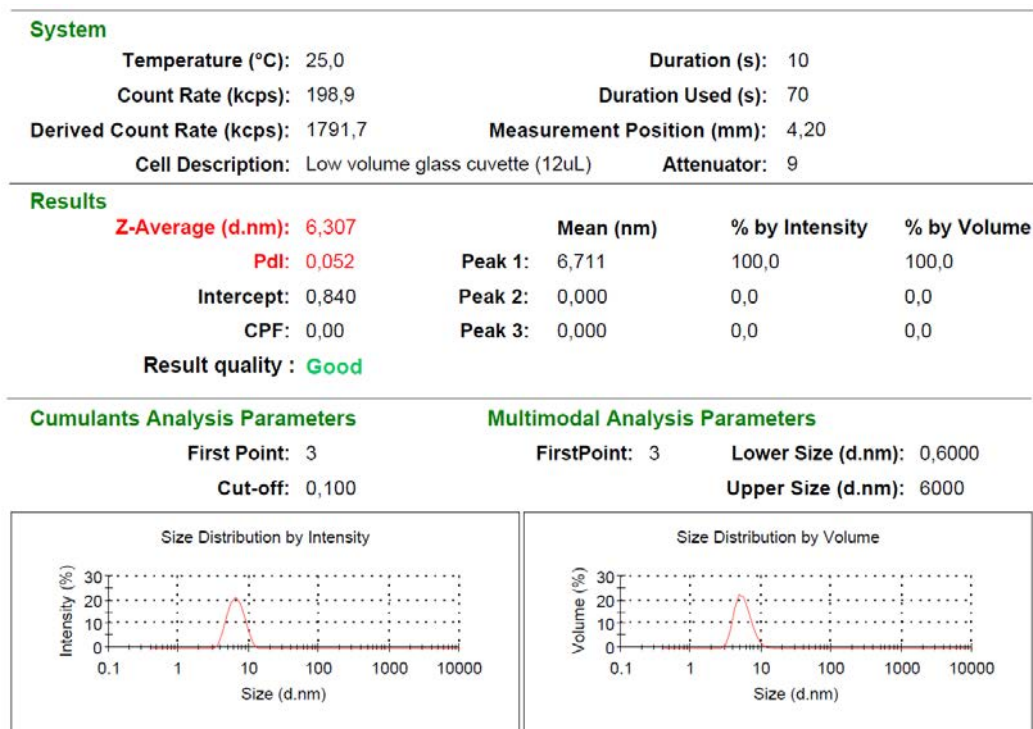


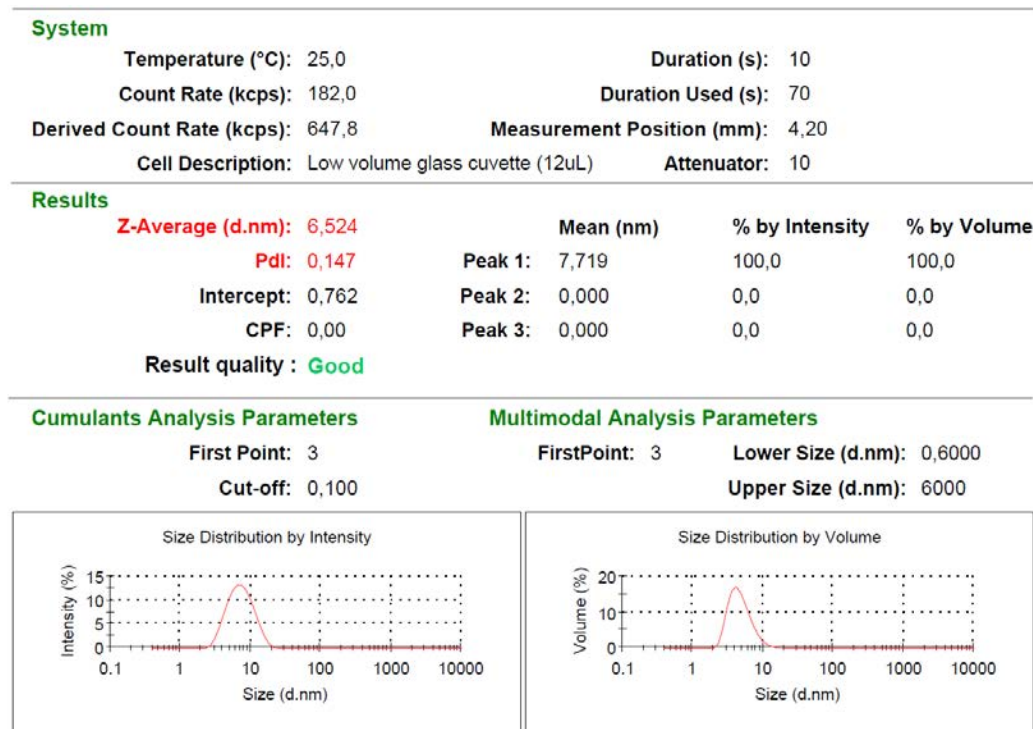
Figure 57 : Gel SDS-PAGE 12,5 % des Lank KR<sub>1</sub> (52,6 kDa), Pik KR<sub>1</sub> (51,4 kDa), Vir KR<sub>6</sub> (47,2 kDa) et Mls KR<sub>A</sub> (50,3 kDa) après purification.

Les quatre protéines ont ensuite été analysées par DLS pour vérifier que les protéines ne s'agrègent pas dans le tampon de stockage ou de filtration su gel. Pour chacune des KR purifiées, l'analyse a montré un indice de polydispersité (PdI) inférieur à 10 % et un seul pic qui indiquent que les protéines ne sont pas agrégées (Figure 58).

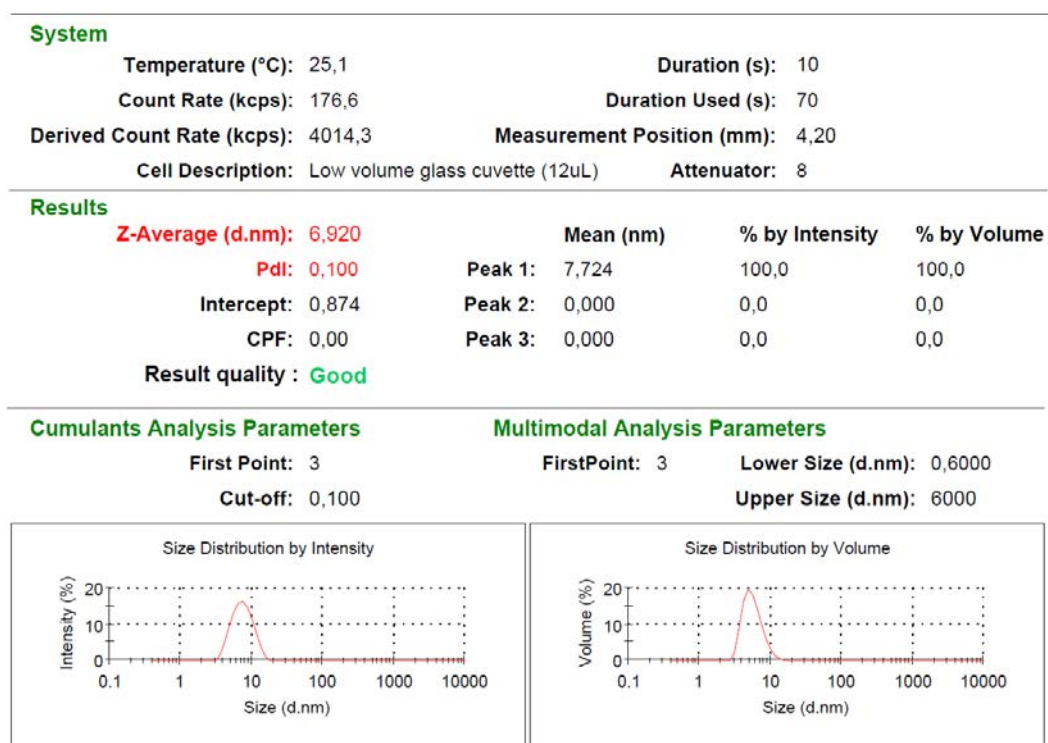
**A**



**B**



C



D

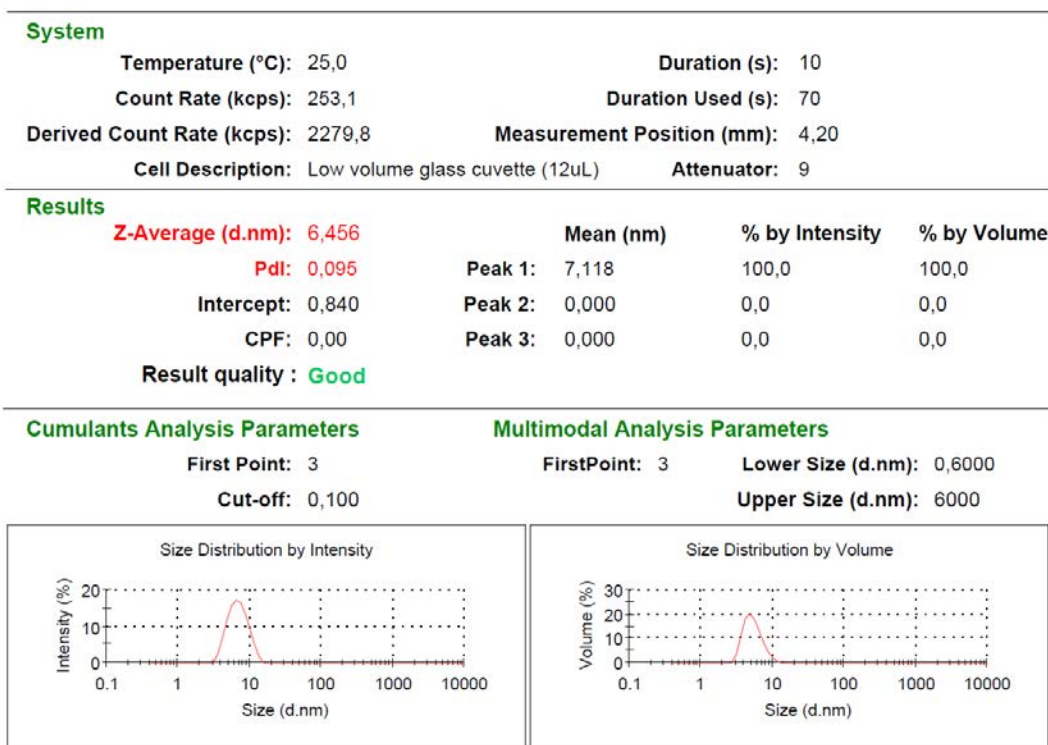


Figure 58 : Données DLS obtenues sur les quatre KR après purification. A : Lank KR<sub>1</sub>. B : Pik KR<sub>1</sub>. C : Vir KR<sub>6</sub>. D : Mls KR<sub>A</sub>.

## II. Cristallisation des KR

### II.1 Etapes de criblage et d'amélioration des premiers cristaux

Les premiers essais de cristallisation en nanogoutte ont été effectués sur les quatre protéines en utilisant les screens PACT premier, JCSG-plus, Morpheus, Wizard, SaltRX, PGA Screen, PEG/Ion Screen, JBS Screen, Index, Crystal Screen et HR. La présence de substrat modèle (*trans*-1-décalone), de cofacteur (NADPH), d'analogue inactif (NADPH<sub>3</sub>), de mélange *trans*-1-décalone/NADPH<sub>3</sub> ou de TCEP (tris(2-carboxyethyl)phosphine) a été testée. La concentration en protéine a été ajustée suivant l'aspect des gouttes de cristallisation. Ainsi, lorsque les gouttes étaient trop claires, la concentration en protéine a été augmentée tandis que lorsque les gouttes contenaient des précipités importants, la concentration en protéine a été diminuée. Pourtant, aucun cristal n'a été obtenu pour les Pik KR<sub>1</sub>, Vir KR<sub>6</sub> et Mls KR<sub>A</sub>. Cependant un cristal était présent pour la Lank KR<sub>1</sub> concentrée à 16 mg/mL auquel est ajouté 2 mM de TCEP dans la condition 2 M de sulfate d'ammonium, 0,1 M de Tris-HCl pH 8,5. La condition a été reproduite en boîte Linbro en faisant varier la concentration en sulfate d'ammonium (agent précipitant) de 1,4 à 2,4 M par incrément de 0,2 M, et en absence et présence d'agent réducteur (TCEP, DTT (dithiothréitol), β-ME (β-mercaptoéthanol)). Des cristaux ont été obtenus pour la protéine avec 2 mM TCEP et à 1,6 M de sulfate d'ammonium avec un mélange de 2 μL de solution de précipitation et 1 μL de protéine (Figure 59).

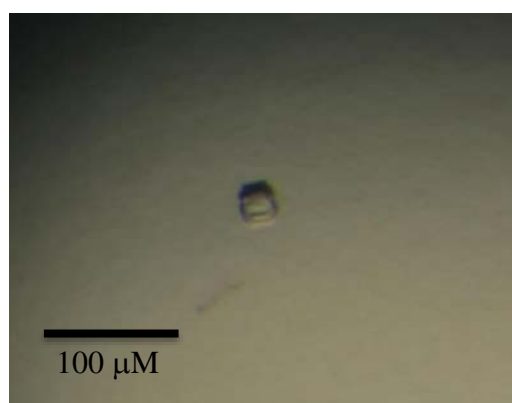


Figure 59 : Premiers cristaux obtenus de la Lank KR<sub>1</sub>.

Les cristaux ont par la suite été améliorés en changeant la valeur du pH ainsi que la nature de la molécule tampon. Les valeurs de pH testées étaient comprises entre

6 et 9 en utilisant du MES (acide 2-(N-morpholino) éthane sulfonique), de l'HEPES (acide 4-(2-hydroxyéthyl)-1- pipérazine éthane sulfonique), du Bis-Tris et du Tris-HCl en tant que molécule tampon à une concentration de 0,1 M. Des cristaux ont été obtenus avec le tampon Tris-HCl pH 8. Après optimisation, il a été possible de réduire le nombre de cristaux par goutte de cristallisation ; cependant tous les cristaux obtenus montraient des macles, c'est-à-dire l'association de plusieurs cristaux orientés différemment. Les plus beaux cristaux obtenus sont montrés en Figure 60. Les spectres de diffractions enregistrés sur la ligne de lumière PROXIMA II du synchrotron SOLEIL montrent un désordre cristallin à basse résolution sur les 74 cristaux que nous avons testés. Nous avons interprété cela comme un désordre cristallin à longue distance, ce qui est cohérent avec la manière dont pousse les cristaux (formation de macle). Il nous a été impossible d'indexer les clichés de diffraction.

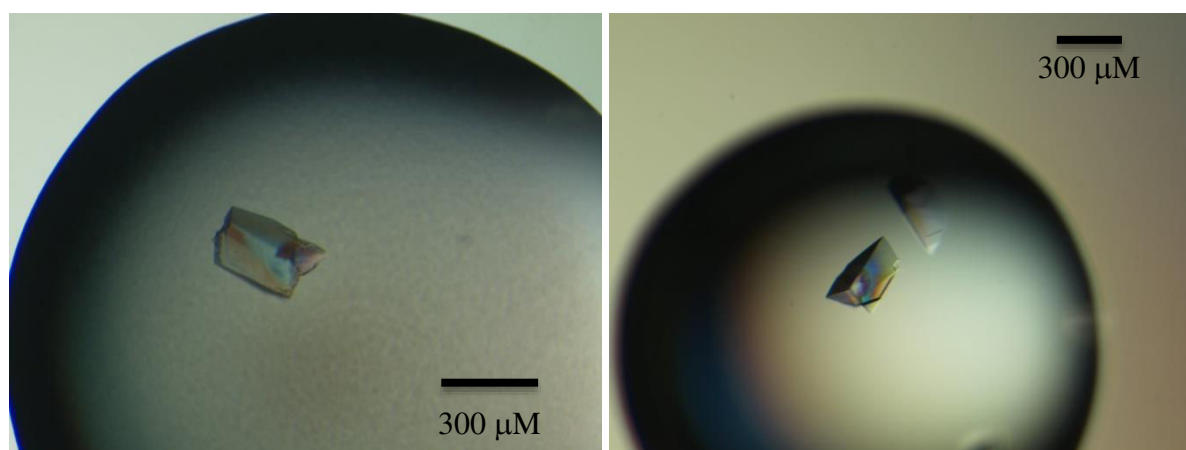


Figure 60 : Cristaux obtenus avec la solution de cristallisation 1,6 M de sulfate d'ammonium, 0,1 M de Tris-HCl pH 8.

## II.2 Utilisation d'additifs

Afin d'empêcher la formation de macles, nous avons réalisé une gamme de sulfate d'ammonium de 1,3 à 1,6 M en ajoutant 0,5 à 10 % de dioxane ou 1 à 10 % de glycérol. La cristallisation a alors été inhibée avec l'obtention de gouttes claires à faible concentration de ces deux molécules et des gouttes précipitées à haute concentration. Nous avons également utilisé le kit des 96 additifs Hampton afin de contrecarrer ce phénomène. Ils ont été ajoutés dans la goutte de cristallisation à des concentrations de 5 et 10 %. Cependant aucun cristal n'a été obtenu dans ces conditions.

### II.3 Micro-ensemencement

Pour tenter d'obtenir des cristaux uniques et non pas associés, nous avons utilisé la technique du micro-ensemencement à partir de cristaux maclés broyés. Trois cristaux maclés de Lank KR<sub>1</sub> ont été lavés plusieurs fois dans la solution de cristallisation afin d'éliminer les précipités. Ils ont été ensuite placés dans 100  $\mu$ L de la solution de cristallisation légèrement plus concentrée en agent précipitant (1,7 M de sulfate d'ammonium) puis broyés à l'aide de billes. Plusieurs dilutions des cristaux broyés ont été effectuées (1/10<sup>ème</sup>, 1/100<sup>ème</sup>, 1/1000<sup>ème</sup>). Les gouttes ont été faites avec un mélange de différents ratios en broyat de cristaux et protéine et ont ensuite été placés au-dessus des puits contenant du sulfate d'ammonium de 0,8 à 1,3 M et 0,1 M Tris-HCl pH 8. Des cristaux ont été obtenus pour les gouttes suspendues au-dessus des puits contenant 1,2 M de sulfate d'ammonium. Les cristaux étant en nombre trop important, ils ont été reproduits avec le broyat dilué au 1/1000<sup>ème</sup>. Les cristaux obtenus étaient de plus petite taille (50–100  $\mu$ m) mais unique sous loupe binoculaire (Figure 61).

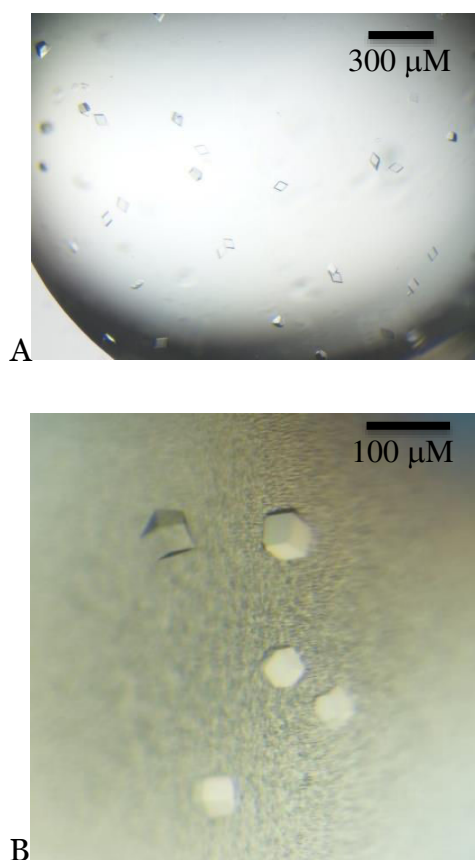


Figure 61 : Cristaux de Lank KR<sub>1</sub> obtenus avec le broyat de cristaux dilués au A : 1/100<sup>ème</sup> et B : 1/1000<sup>ème</sup>.

Ces cristaux ont été congelés sous flux d'azote, puis stockés en attendant la mission synchrotron.

### III. Enregistrement et traitement des données diffraction

Plusieurs jeux de données ont été enregistrés sur différents cristaux de la Lank KR<sub>1</sub> sur la ligne PROXIMA I du synchrotron SOLEIL. Les clichés ont été enregistrés sur 360° avec une image par degré. Les données ont été indexées puis intégrées avec XDS<sup>92</sup>. XSCALE a été utilisé pour la mise à l'échelle<sup>92</sup>. Un exemple de la qualité d'un jeu de données est représenté dans le Tableau 4. Les cristaux contiennent 40% de solvant pour un coefficient de Matthews de 2,08 Å<sup>3</sup>/Da<sup>101</sup>.

Groupe d'espace	P2 <sub>1</sub>
Paramètres de maille a, b, c (Å)	69,7 47,3 70,6
$\alpha, \beta, \gamma$ (°)	90,0 118,3 90,0
Résolution (Å)	50-1,90
I/ $\sigma$ (I)	13,51 (2,54)
R <sub>merge</sub> (%)	8,2 (67,9)
CC (1/2) (%)	99,8 (89,3)
Complétude (%)	95,8 (85,2)

Tableau 4 : Qualité des données de diffraction après mise à l'échelle.

## IV. Phasage

### IV.1 Remplacement moléculaire

Les structures des domaines KR<sub>1</sub> de DEBS<sup>43</sup>, KR<sub>1</sub> de la PKS tylosine<sup>60</sup>, KR<sub>2</sub> de la PKS amphotéricine<sup>102</sup>, KR<sub>3</sub> de la PKS pikromycine<sup>103</sup> et KR<sub>1</sub> de la PKS phoslactomycine<sup>65</sup> ont déjà été résolues et leurs séquences partagent environ 60 % d'identité avec celle de la Lank KR<sub>1</sub>.

Dans ce contexte, le remplacement moléculaire est une méthode de choix pour le phasage. Les coordonnées h, k et l ainsi que le fichier PDB de la structure d'un domaine homologue ont été utilisés par les logiciels Phaser<sup>93</sup> et MolRep<sup>94</sup>. En utilisant ces deux programmes, la fonction de rotation est contrastée ; par contre, la fonction de translation ne l'est pas. Ceci est typique d'une erreur dans le choix du groupe d'espace. Les données ont donc été retraitées et un grand soin a été porté à la détermination de

la maille et du réseau cristallin notamment en termes de réflexions indexées et rejetées. Les deux groupes d'espace possibles de plus hautes symétries, monoclinique centré et orthorhombique centré, montrent une mauvaise qualité d'ajustement lors de l'indexation. De plus, les facteurs  $R_{\text{merge}}$  avoisinent les 40% après intégration. L'existence d'une pseudo symétrie a donc été émise. Les données ont donc aussi été traitées en P1. Que ce soit en P1 ou en P2<sub>1</sub>, la fonction de translation n'est pas contrastée et les solutions obtenues montrent de nombreux « clashes » stériques entre symétriques (protéines liées par symétrie), quels que soit les modèles utilisés. Nous avons également utilisé un modèle qui contient uniquement les structures secondaires pour s'affranchir des éléments mobiles tels que les boucles, mais sans succès. Il semble donc que le faible pourcentage de solvant soit responsable de l'exclusion de la bonne solution lors de l'étape de translation. Nous avons aussi joué sur le diamètre d'intégration, sans succès.

## IV.2 Obtention de dérivés lourds

D'autres méthodes existent pour résoudre le problème de la phase, par exemple le SAD ou le MAD. Ces méthodes utilisent le signal anomal d'un élément présent dans le cristal. Une des méthodes les plus utilisées est de produire une protéine sélénée. Cependant le domaine KR de Lank KR<sub>1</sub> possède deux méthionines pour 488 résidus alors qu'un minimum d'une méthionine pour 100 résidus est préconisé. Nous avons donc décidé d'écarter cette approche et d'utiliser plutôt les lanthanides afin d'essayer de phaser. Pour obtenir des cristaux en présence de ces éléments (Gd, Eu, Yb), nous avons soit co-cristallisé la protéine en présence des différents lanthanides, soit trempé les cristaux dans une solution contenant l'atome lourd. Des cristaux ont été obtenus par co-cristallisation en présence de 10 mM de Gd, Eu-DOTA, Yb-DOTA, Eu-HPDO<sub>3</sub>A et Yb-HPDO<sub>3</sub>A.

Les cristaux ont ensuite été trempés dans la solution de cristallisation afin d'éliminer les traces de lanthanides à la surface des cristaux puis congelés. Sur la ligne Proxima I, les spectres de fluorescence n'ont pas mis en évidence la présence d'Yb, d'Eu et de Gd dans les cristaux. Toutefois un jeu de données a été enregistré à partir des cristaux obtenus par co-cristallisation en présence de Gd. Après intégration, aucun signal anomal n'a pu être mis en évidence.



Les tentatives pour obtenir des cristaux de dérivés lourds plus classiques, en présence de platine (Pt) ou mercure (Hg) par exemple, se sont aussi avérées infructueuses par trempage. En effet, les spectres de fluorescence n'ont pas mis en évidence la présence de Pt ou Hg dans les cristaux. La présence de sels d'atome lourds « classiques » dans les conditions de cristallisations conduit à l'obtention de cristaux inorganiques en raison de la présence de TCEP dans les conditions de cristallisation.

## Partie 4 : Etude de la stabilisation de la stéréochimie du C $\alpha$

### I. Clonage et purification de PIKS ACP<sub>1</sub>, KS<sub>1</sub>-AT<sub>1</sub> et KR<sub>1</sub>

#### I.1 Clonage et purification du domaine ACP<sub>1</sub>

Pour ce projet, tous les domaines du module 1 de la PKS pikromycine ont été sous-clonés et purifiés : le domaine ACP<sub>1</sub>, le didomaine KS<sub>1</sub>-AT<sub>1</sub> et le domaine KR<sub>1</sub>. Le gène codant l'ACP<sub>1</sub> est synthétique et a été commandé chez DNA 2.0. Il a été sous-cloné dans le vecteur pET28b par le Dr C. Jacob. La séquence optimisée pour une expression chez *E. coli* de l'ACP<sub>1</sub> ainsi que la méthode de clonage sont basées sur la publication de Valenzano<sup>66</sup>. Le vecteur pET28b-ACP<sub>1</sub> a ensuite été utilisé pour transformer des cellules BL21 (DE3) d'*E. coli* rendues compétentes. La production de la protéine est réalisée en milieu LB kanamycine pour la protéine non marquée ou en milieu M9 kanamycine supplémenté en <sup>15</sup>NH<sub>4</sub>Cl avec et sans glucose <sup>13</sup>C pour la protéine marquée isotopiquement <sup>13</sup>C/<sup>15</sup>N et <sup>15</sup>N respectivement. Les cellules ont été lysées par sonication, et après centrifugation le surnageant a été chargé sur une colonne de Nickel-Sépharose. Les fractions éluées contenant l'ACP<sub>1</sub> ont été assemblées et dialysées sur la nuit en présence de thrombine afin de cliver le tag histidine. Ensuite, cet échantillon a été injecté à nouveau sur une colonne de Nickel-Sépharose et la fraction non-retenue contenant la protéine a été collectée et injectée sur une colonne de filtration sur gel. L'ACP<sub>1</sub> ne possède aucun résidu aromatique, les différentes fractions d'éluition ont donc été déposées sur gel SDS-PAGE 12,5% afin d'identifier celles où la protéine était présente. La protéine obtenue est pure (Figure 62). La concentration en protéine a ensuite été mesurée par la méthode du dosage selon Bradford. Une courbe étalon a été réalisée grâce à une gamme de concentration de BSA (albumine de sérum bovin) (Figure 63) qui nous a permis de déterminer la concentration en protéines et donc les rendements de purification. Ces derniers sont de 2 mg d'ACP<sub>1</sub> non marqué/L de culture, 1,45 mg d'ACP<sub>1</sub> <sup>15</sup>N/L de culture et de 0,72 mg d'ACP<sub>1</sub> <sup>13</sup>C <sup>15</sup>N/L de culture. Les solutions protéiques non concentrées ont ensuite été congelées dans l'azote liquide puis conservées à -80°C. L'ACP<sub>1</sub> <sup>15</sup>N et l'ACP<sub>1</sub> <sup>13</sup>C <sup>15</sup>N ont été purifiés par A. Naretto durant son stage de M1 co-encadrée par le Dr C. Jacob et moi-même.

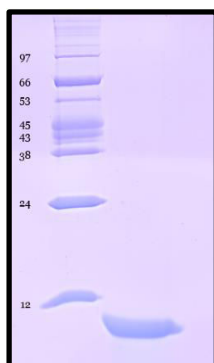


Figure 62 : Gel SDS-PAGE 12,5 % de l'ACP<sub>1</sub> (10,6 kDa) non marqué après purification.

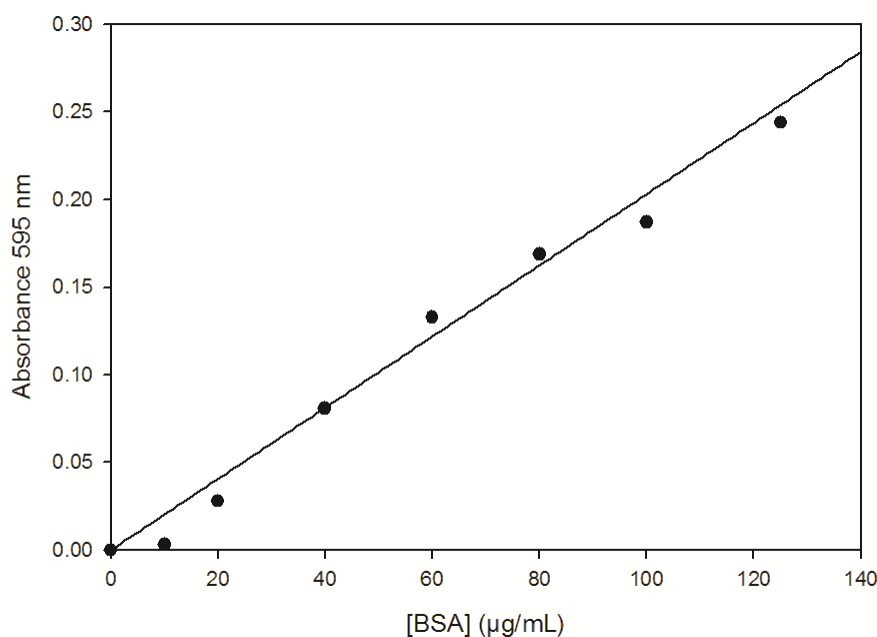


Figure 63 : Courbe étalon du dosage protéique selon la méthode de Bradford.

## I.2 Clonage et purification du didomaine KS<sub>1</sub>-AT<sub>1</sub>

Le clonage est réalisé par ligation du gène synthétique codant le didomaine KS<sub>1</sub>-AT<sub>1</sub> dans le vecteur pET28b. Il a été effectué par le Dr C. Jacob. Le protocole est ensuite similaire à celui de la purification du domaine ACP<sub>1</sub>. Le vecteur a été utilisé pour transformer dans des cellules compétentes BL21 (DE3) d'*E. coli*. La production de la protéine s'est faite en milieu LB supplémenté en kanamycine. Les cellules ont été lysées puis le surnageant a été centrifugé. La protéine a été purifiée sur une colonne de Nickel-Sépharose. L'éluat a ensuite été dialysé en présence de thrombine afin de cliver le tag histidine. Cette étape a été suivie par une seconde colonne de Nickel-Sépharose. Les fractions non-retenues par la colonne (contenant la protéine) ont été concentrées et injectées sur une colonne de filtration sur gel. La pureté de l'échantillon protéique a été

contrôlée par gel SDS-PAGE 12,5% (Figure 64). La protéine a été concentrée et congelée dans l'azote liquide pour être stockée à  $-80\text{ }^{\circ}\text{C}$ . Le rendement de purification est de 64 mg de protéine/L de culture.

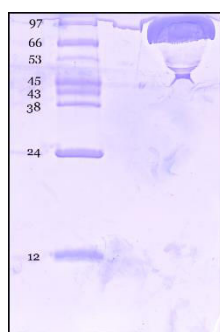


Figure 64 : Gel SDS-PAGE 12,5 % du didomaine KS<sub>1</sub>-AT<sub>1</sub> (98,2 kDa) après purification.

### I.3 Clonage et purification du domaine KR<sub>1</sub>

La séquence codant la KR<sub>1</sub> a été amplifiée à partir de l'ADN génomique de *S. venezuelae*. Elle a été clonée dans le vecteur pBG106. Le vecteur a été utilisé pour transformer dans des cellules compétentes BL21 (DE3) d'*E. coli*. La production protéique a été réalisée en milieu ZYM 5052. Les cellules ont ensuite été lysées et le surnageant a été purifié en plusieurs étapes. La première étape a consisté en une colonne de Nickel-Sépharose. L'échantillon protéique obtenu a ensuite été dialysé en présence de protéase PreScission afin de cliver le tag histidine, puis déposé sur une nouvelle colonne de Nickel-Sépharose. La purification s'est terminée par une colonne de filtration sur gel. La pureté a été contrôlée par gel SDS-PAGE 12,5% (Figure 65). La protéine a été congelée dans l'azote liquide et stockée à  $-80\text{ }^{\circ}\text{C}$ . Le rendement de purification est de 7 mg de KR<sub>1</sub>/L de culture.

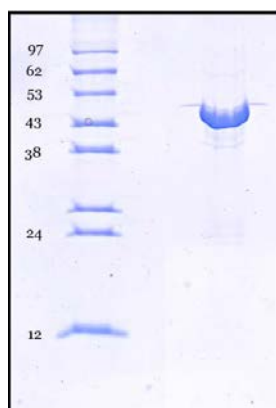


Figure 65 : Gel SDS-PAGE 12,5 % de la KR<sub>1</sub> (51,4 kDa) après purification.

## II. Attribution de la chaîne principale de l'ACP<sub>1</sub>

Afin d'observer par RMN une éventuelle interaction entre le domaine ACP<sub>1</sub> et le dikétide, il était indispensable d'attribuer les déplacements chimiques du spectre <sup>1</sup>H-<sup>15</sup>N HSQC de l'ACP<sub>1</sub> (c'est-à-dire, d'attribuer les NH de la chaîne principale). En effet, la comparaison des pics des spectres <sup>1</sup>H-<sup>15</sup>N HSQC de l'ACP<sub>1</sub> *apo*, *holo* ou acylé avec le dikétide permet d'observer des variations de déplacements chimiques dues à des éventuelles interactions entre l'ACP<sub>1</sub> et le bras phosphopantéthéine ou le dikétide.

Pour faire ces attributions, l'ACP<sub>1</sub> doublement marqué et sous forme *apo* a été concentré à 1 mM. Les déplacements chimiques des spectres <sup>1</sup>H-<sup>15</sup>N HSQC, HNCACB, HNCA, HNCO, HNHA et CBCACONH ont été enregistrés et attribués grâce au logiciel Sparky par le Dr. B. Chagot.

L'ensemble des fonctions NH de la protéine a été attribué, mises à part celle de la sérine portant le bras phosphopantéthéine, qui ne présentait pas de pic au niveau du spectre <sup>1</sup>H-<sup>15</sup>N HSQC, mais également celles des Gly16, Arg18 et Thr19 (Figure 66). Cette région a pu être manquante du fait du pH non optimal d'enregistrement (7,5). Le proton amide de la sérine est sans doute échangé avec le solvant et est donc invisible.

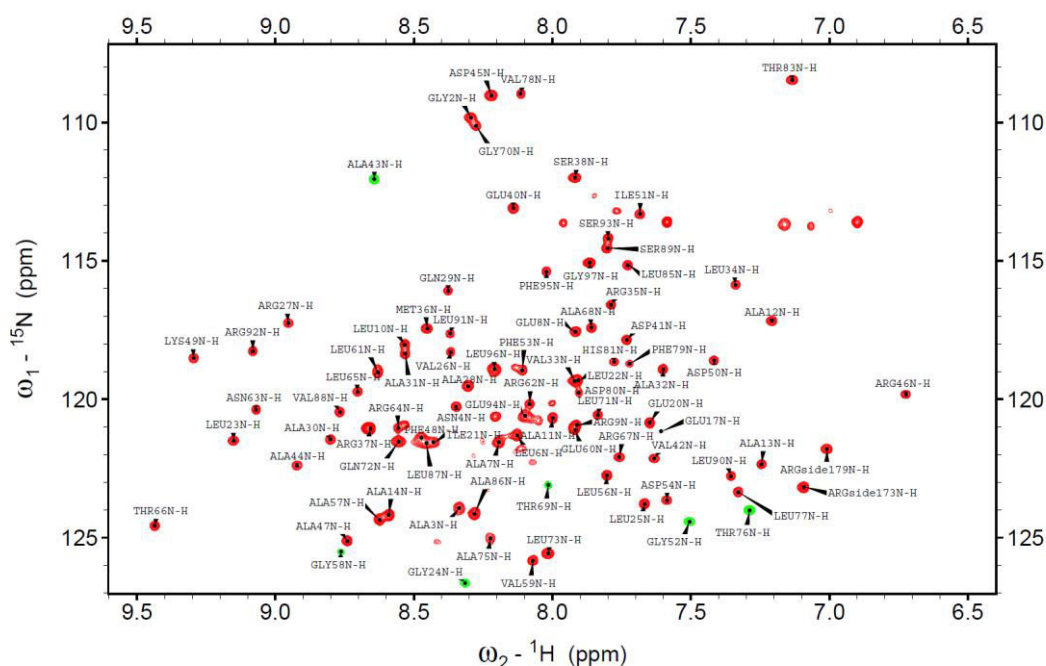


Figure 66 : Spectre <sup>1</sup>H-<sup>15</sup>N HSQC de l'ACP<sub>1</sub> *apo* attribué.

### III. Formation de la forme *holo* de l'ACP<sub>1</sub> et analyse de la modification

#### III.1 Réaction de modification

Dans le but de modifier le domaine ACP<sub>1</sub>, deux phosphopantéthéinyl transférases (PPTases) différentes ont été utilisées, la Sfp et la Svp, afin d'observer si une des deux était plus réactive. La Sfp a été purifiée par le Dr. B. Chagot tandis que la Svp par A. Naretto. Le milieu réactionnel de modification était identique pour les deux PPTases : 100 µM d'ACP ont été mélangés à 3 µM de Sfp ou Svp, 5 mM de CoA, 10 mM de MgCl<sub>2</sub>, 5 mM de DTT, et complétés à 300 µL avec du tampon phosphate 100 mM, pH 7,5. La réaction s'est déroulée sur la nuit à température ambiante. Enfin, l'ACP<sub>1</sub> a été purifié par filtration sur gel pour le séparer du CoA et de la Sfp ou de la Svp.

Afin de vérifier si la réaction de modification a eu lieu, des expériences de spectrométrie de masse ont été réalisées sur le produit des deux réactions.

#### III.2 Confirmation de la modification par spectrométrie de masse

En collaboration avec le Dr. H. Mazon, l'ajout du bras phosphopantéthéine a été confirmé par une expérience ESI-Q-TOF en condition dénaturante. Les masses de l'ACP<sub>1</sub> *apo* ainsi que l'ACP<sub>1</sub> modifié par la Sfp ont été analysées. Le spectre de masse de l'ACP<sub>1</sub> *apo* a présenté un pic à 10600,8 Da correspondant bien à la masse attendue de la protéine (masse théorique = 10601,2 Da). Quant à l'ACP<sub>1</sub> qui a été modifié, le spectre de masse a présenté un pic à 10941,0 Da. L'ACP<sub>1</sub> modifié présente donc une augmentation de masse de 340,2 Da. Sachant que le bras phosphopantéthéine possède une masse de 340,3 Da et donc que la protéine *holo* a une masse théorique de 10941,5 Da, nous avons donc pu confirmer que l'ACP<sub>1</sub> était bien sous forme *holo* (Figure 67)

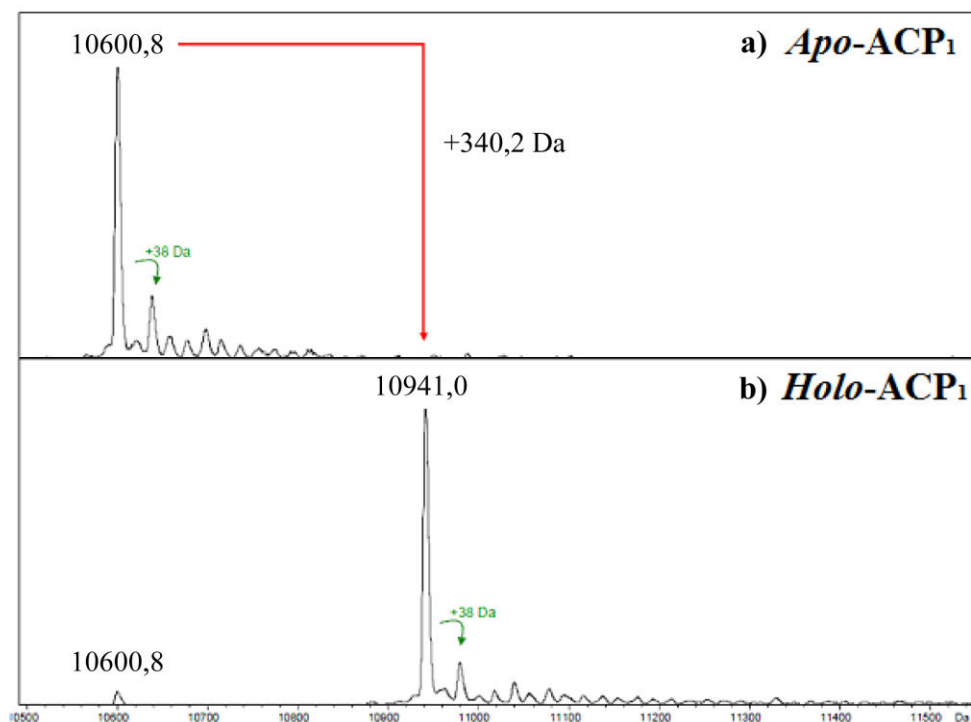


Figure 67 : Spectres de masse déconvolués de l'ACP<sub>1</sub> apo (a) et holo (b) modifié par la Sfp. L'augmentation de masse entre les deux formes correspond à la masse du bras phosphopantéthéine (340,3 Da).

La même expérience a été réalisée avec l'ACP<sub>1</sub> modifié par la Svp. Les spectres de masse de l'ACP<sub>1</sub> apo et modifié ont été analysés. Le spectre de l'ACP<sub>1</sub> apo a présenté un pic de masse à 10600,7 Da alors que celui de l'ACP<sub>1</sub> modifié a présenté un pic à 10941,2 Da. Il y a donc une augmentation de masse de 340,5 Da, confirmant que la modification de l'ACP<sub>1</sub> a bien été réalisée (Figure 68).

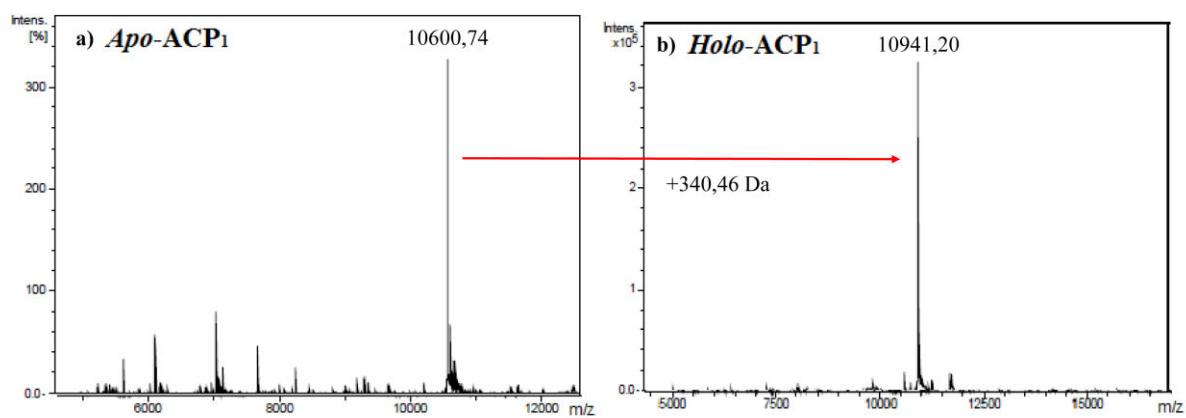


Figure 68 : Spectres de masse déconvolués de l'ACP<sub>1</sub> apo (a) et holo (b) modifié par la Svp. L'augmentation de masse entre les deux formes correspond à la masse du bras phosphopantéthéine.

### III.3 Analyse de la position de modification de l'ACP<sub>1</sub>

Dans le même temps, une autre technique de spectrométrie de masse a été utilisée afin de vérifier si l'ajout du bras phosphopantéthéine a eu lieu sur la sérine conservée. Cette vérification qui pourrait se faire par RMN, *via* l'analyse du déplacement chimique de la fonction NH de la sérine, est impossible puisque celle-ci n'est pas visible. Des études de séquençage par technique d'empreinte peptidique ont alors été réalisées sur l'ACP<sub>1</sub> *apo* et *holo* en collaboration avec le Dr J. B. Vincourt. Au préalable, les protéines ont été digérées à la trypsine dans le but de générer des peptides qui ont ensuite été analysés. Des différences sont visibles entre les spectres provenant des digestions de l'ACP<sub>1</sub> *apo* et l'ACP<sub>1</sub> modifié par la Sfp et ou la Svp. Un pic avec une masse de 1737.9 Da est présent pour le spectre de l'ACP<sub>1</sub> *apo* et absent pour les deux spectres des ACP<sub>1</sub> *holo*, tandis qu'un pic avec une masse de 2135 Da est présent pour les deux spectres des ACP<sub>1</sub> *holo* mais absent pour le spectre de l'ACP<sub>1</sub> *apo* (Figure 69). La différence de masse de 397,1 Da est supérieure à la masse du bras. Ceci est dû à l'étape de carbamidométhylation réalisée dans le but de bloquer les fonctions thiols des cystéines avant la protéolyse. Cette réaction s'effectue également sur la fonction thiol de la phosphopantéthéine, entraînant cette augmentation de masse de 57 Da. Cette différence de pic, observée entre les spectres de masse de l'ACP<sub>1</sub> *apo* et *holo* tend donc à suggérer que ce peptide est celui qui porte le bras phosphopantéthéine.



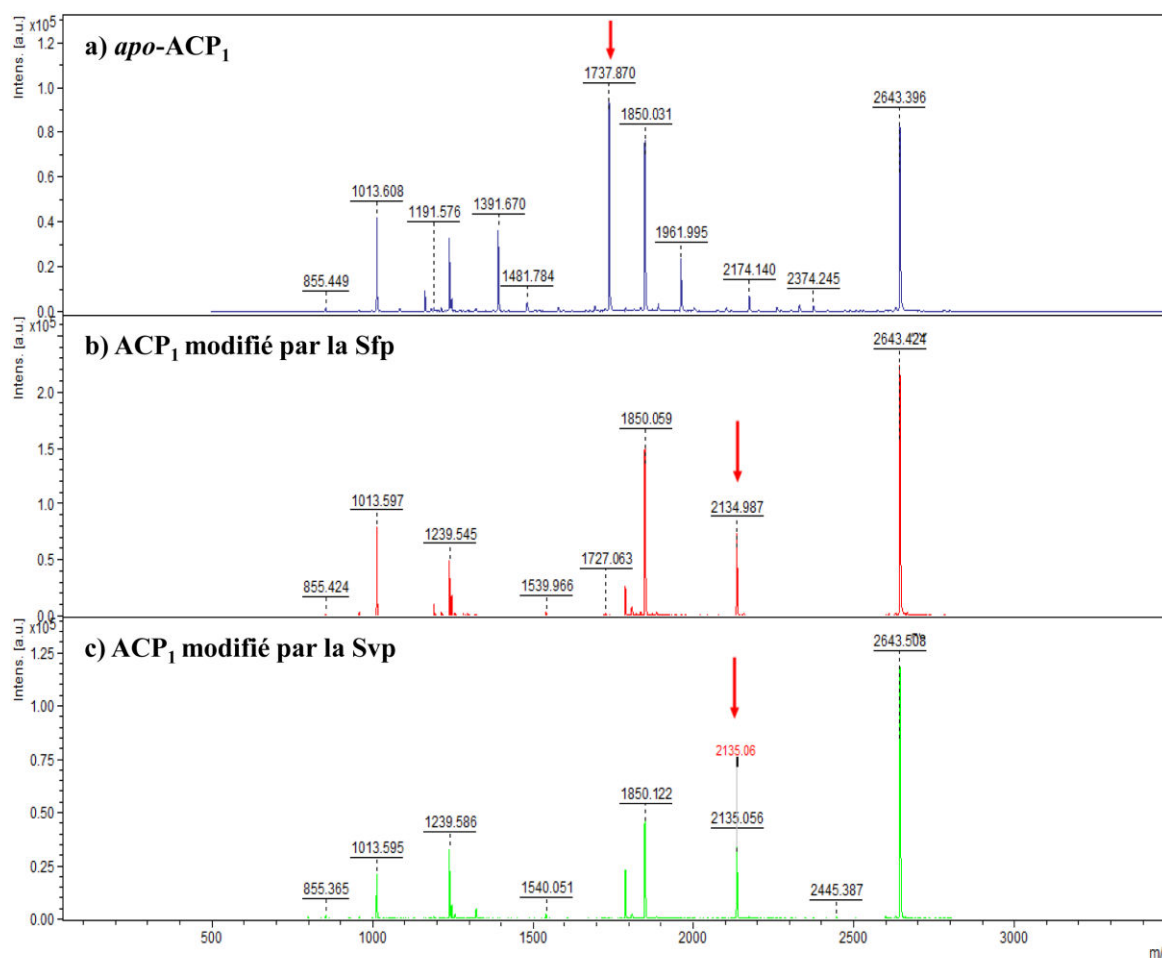


Figure 69 : Séquençage par technique d’empreinte de l’ACP<sub>1</sub> apo (a) et holo modifiés par la Sfp (b) et la Svp (c).

De plus, le logiciel Mascot qui permet d’identifier la séquence d’un peptide correspondant à un pic de masse, a reconnu la séquence suivante comme étant celle modifiée : AFKDIGFDSL~~G~~VELR. Ce résultat est confirmé par la présence dans cette séquence du motif consensus DSL qui entre en jeu dans la reconnaissance ACP/PPTase et porte la sérine (S) sur laquelle la modification par la phosphopantéthéine a lieu<sup>104</sup>.

L’identification du résidu modifié est confirmée par le fait que le peptide a été fragmenté. Ceci permet ainsi de reconstituer sa séquence grâce aux masses des différents fragments engendrés (Figure 70).

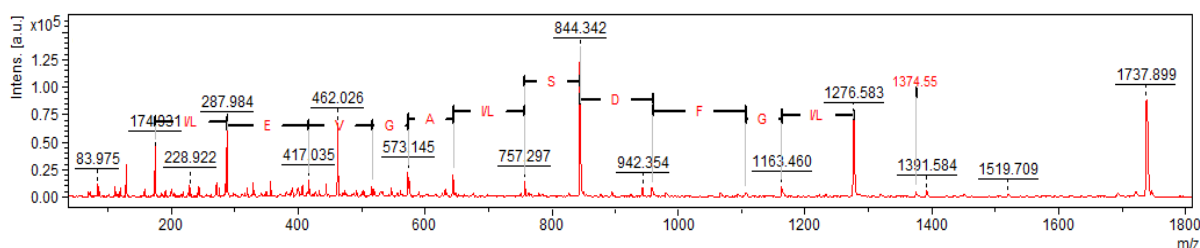


Figure 70 : Reconstitution de la séquence du peptide modifié par spectrométrie de masse après fragmentation.

### III.4 Analyse des modifications par RMN en utilisant la Sfp

A partir de l'analyse des modifications par spectrométrie de masse, seule la Sfp a été utilisée pour modifier l'ACP<sub>1</sub>, puisqu'aucune différence n'a été observée entre la modification par la Sfp et celle par la Svp.

Ainsi, l'ACP<sub>1</sub> <sup>15</sup>N a été convertie en forme *holo* par la Sfp suivant le protocole décrit précédemment, puis la protéine a été purifiée par un passage sur une colonne de filtration sur gel (séparation du CoA, de la Sfp et de l'ACP<sub>1</sub>) et concentrée à 100 μM. Un spectre <sup>1</sup>H-<sup>15</sup>N HSQC a ensuite été enregistré en utilisant l'ACP<sub>1</sub> *holo* ainsi obtenu. Un autre spectre <sup>1</sup>H-<sup>15</sup>N HSQC a également été enregistré avec de l'ACP<sub>1</sub> *apo*. La différence de déplacements chimiques entre les deux spectres a été quantifiée de manière à distinguer les résidus dont l'environnement a changé sous l'effet de la modification avec le bras phosphopantéthéine (Figure 71). Plusieurs résidus sont affectés par la modification : les résidus Ile21, Phe53, asp54, Leu56, Gly58, Val59, Glu60, Thr66, Gly 70, Leu73, Thr76, Leu77, Val78, Phe79, Asp80 et His81. Ces résidus subissent un changement au niveau de leur déplacement chimique, probablement provoqué par une perturbation de l'environnement due à l'ajout du bras. Un modèle de l'ACP<sub>1</sub> est généré par I-TASSER afin de placer ces résidus dans la structure<sup>105</sup>. Ces résidus se situent à proximité de la sérine conservée et donc à proximité du bras (Figure 72). La sérine conservée de la phosphopantéthéine n'est pas visible mais elle aurait dû présenter un décalage du déplacement chimique.

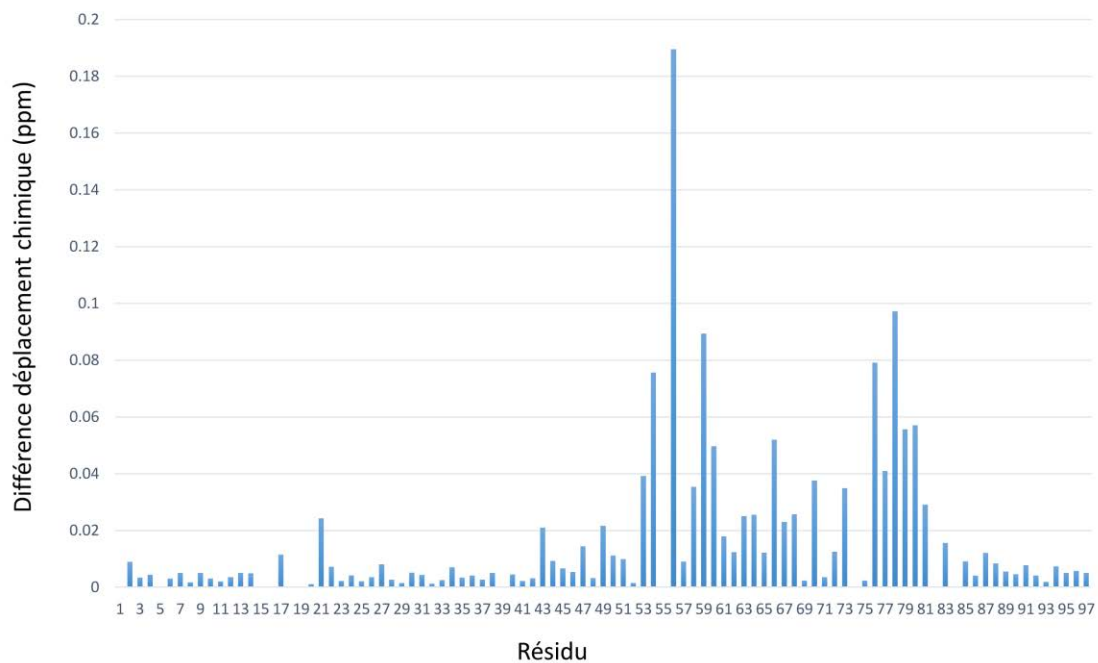


Figure 71 : Différences de déplacement chimique entre l'ACP<sub>1</sub> apo et holo pour chaque résidu.

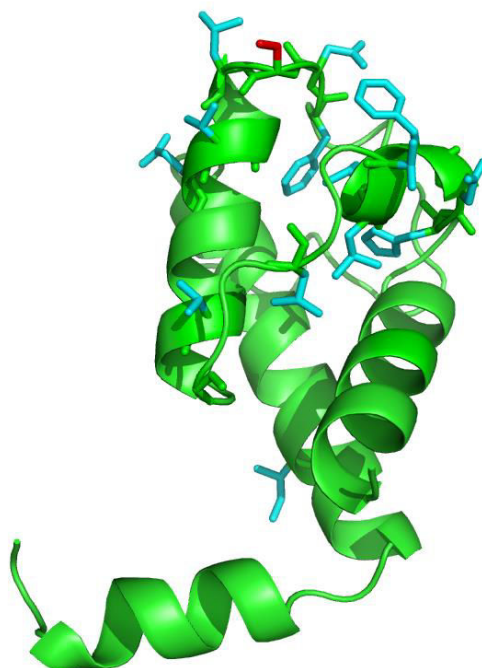


Figure 72 : Modèle de la structure de l'ACP<sub>1</sub>. Les résidus en bleu subissent une modification de leur déplacement chimique après la modification. Ils se situent à proximité de la sérine conservée en rouge.

#### IV. Acylation de l'ACP<sub>1</sub> par le CoA-dikétide et analyse de la modification

Ayant établi les conditions pour la modification efficace de l'ACP<sub>1</sub> avec le bras phosphopantéthéine, la prochaine étape sera de le dériver avec du dikétide qui est l'intermédiaire naturel généré par la réaction de condensation au sein du module 1. Cette réaction se fera à partir de dikétide-CoA (produit par voie enzymatique à partir de dikétide-pantéthéine synthétique) en employant la Sfp comme PPTase. Celle-ci va permettre, par la suite, de voir s'il y a des interactions entre le dikétide et l'ACP<sub>1</sub> et donc si l'ACP<sub>1</sub> est bien capable de stabiliser la configuration du dikétide comme proposé précédemment<sup>66,72</sup>. Dans un premier temps, l'ACP<sub>1</sub> <sup>15</sup>N sera modifié par du dikétide <sup>12</sup>C et un spectre <sup>1</sup>H-<sup>15</sup>N HSQC sera enregistré. Ce dernier sera comparé avec les spectres <sup>1</sup>H-<sup>15</sup>N HSQC des ACP<sub>1</sub> *apo* et *holo* déjà acquis afin d'observer s'il y a des changements dans les déplacements chimiques qui résulteraient d'une interaction entre le dikétide et l'ACP<sub>1</sub>. Dans un second temps, c'est une étude avec le dikétide <sup>13</sup>C qui permettra de confirmer les conclusions. Des spectres 1D carbone et COCA seront enregistrés avec la pantéthéine-diketide <sup>13</sup>C et CoA-dikétide <sup>13</sup>C seuls et également avec l'ACP<sub>1</sub> <sup>12</sup>C <sup>14</sup>N modifié avec le dikétide <sup>13</sup>C afin d'observer s'il y a des changements dans les déplacements chimiques du dikétide qui sont cohérents avec une interaction dikétide/ACP<sub>1</sub> (Figure 73).

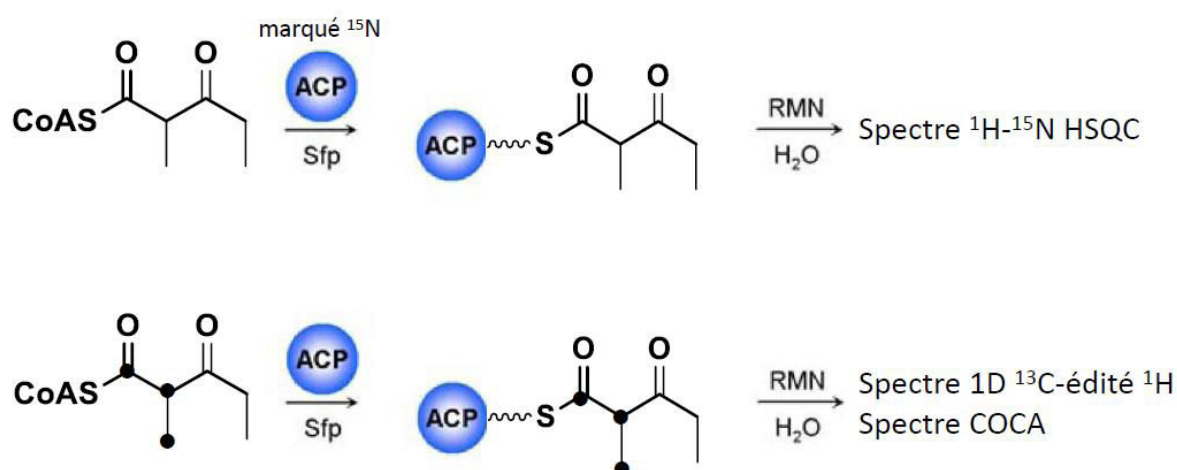


Figure 73 : Expériences à réaliser dans le but de déterminer si la stabilisation du dikétide est induite par son interaction avec l'ACP<sub>1</sub>.

## IV.1 Synthèse des substrats

### IV.1.1 Purification des enzymes de synthèse du substrat

La synthèse du substrat CoA-dikétide, qu'il soit isotopiquement marqué ou non, est réalisée par les trois enzymes PanK (pantothénate kinase), PPAT (phosphopantéthéine adényltransférase) et DPCK (déphospho-CoA kinase)<sup>90</sup>. Les plasmides codant les trois protéines ont été fournis par le Pr. P. F. Leadlay (Université de Cambridge, UK). Les plasmides ont été utilisés pour transformer des cellules compétentes BL21(DE3) d'*E. coli*. La production des enzymes est réalisée en milieu LB supplémenté en kanamycine pour PanK et PPAT et en ampicilline pour DPCK. Les cellules ont été lysées par sonication puis le surnageant centrifugé a été chargé sur une colonne de Nickel-Sépharose. Les fractions éluées contenant la protéine d'intérêt ont été déposées sur gel SDS-PAGE 12,5 % et assemblées (Figure 74). Le tout a été concentré et congelé dans l'azote liquide puis stocké à  $-80^{\circ}\text{C}$ . Les rendements de purification sont de 60 mg de PanK /L de culture, 40 mg de PPAT /L de culture et 45 mg de DPCK /L de culture.

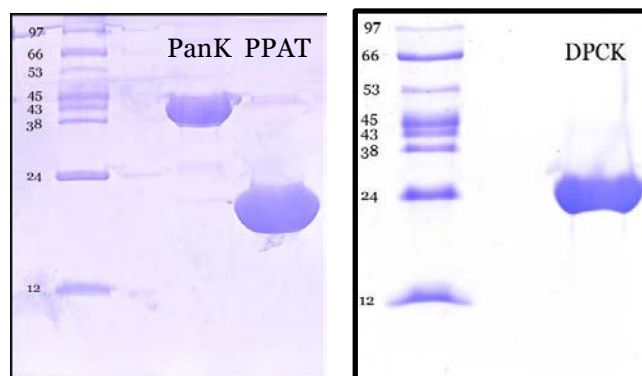


Figure 74 : Gel SDS-PAGE 12,5 % de PanK (36,4 kDa), PPAT (17,6 kDa) et DPCK (23,9 kDa).

### IV.1.2 Synthèse du substrat

La phosphopantéthéine-dikétide  $^{12}\text{C}$  ou  $^{13}\text{C}$  (synthétisé par l'équipe du Pr. A. Kirschning, Université de Hannovre) a été mis en solution avec un mélange des trois protéines de modification (PanK, PPAT et DPCK) ainsi que de l'ATP. La réaction s'est déroulée pendant 3 h à  $30^{\circ}\text{C}$ , puis le mélange a été lyophilisé. Le substrat (13 mM) a été repris dans un mélange eau 95 %/acétonitrile 5% en supposant que la réaction a été totale au vu de la quantité de substrat ajoutée au début de la réaction.

### IV.1.3 Analyse par spectrométrie de masse de l'obtention des CoA-dikétides

Afin de vérifier que la réaction de synthèse s'est correctement déroulée et qu'elle est totale, des expériences de spectrométrie de masse ont été réalisées sur les produits de la synthèse du CoA-dikétide  $^{12}\text{C}$  et CoA-dikétide  $^{13}\text{C}$ , en collaboration avec C. Paris (LIBio, Université de Lorraine). Concernant le CoA-dikétide  $^{12}\text{C}$ , la chromatographie en phase liquide donne un pic qui présente un temps de rétention de 2,71 min. L'entité présente à ce temps de rétention possède un spectre de masse avec un pic de masse majoritaire à 880,2 Da correspondant à la masse du CoA-dikétide. (Figure 75).

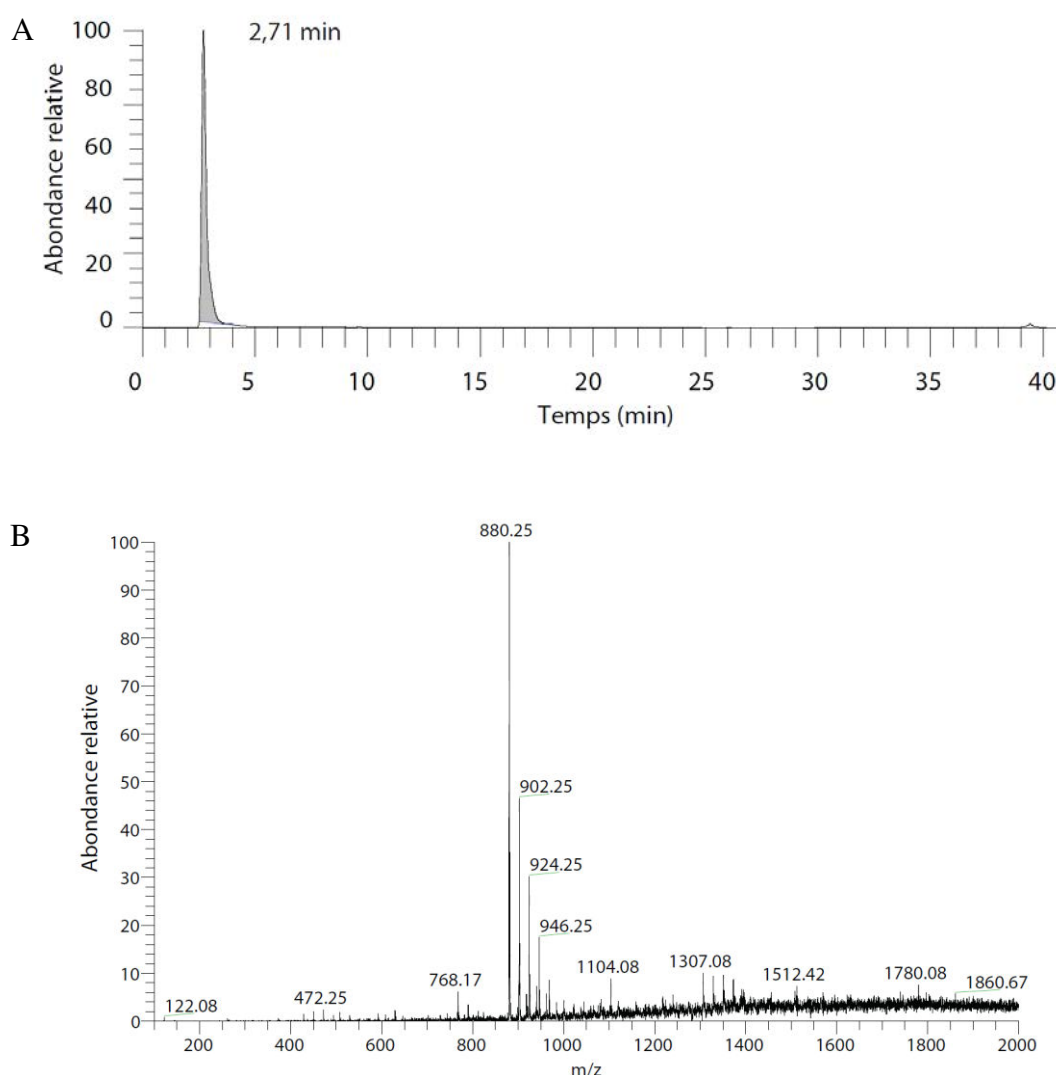


Figure 75 : Analyse par LC-MS du CoA-dikétide  $^{12}\text{C}$  synthétisé. A : partie chromatographie en phase liquide. B : spectrométrie de masse.

Pour le CoA-dikétide  $^{13}\text{C}$ , la chromatographie en phase liquide donne un pic qui présente un temps de rétention de 2,71 min, c'est-à-dire précisément au même temps de rétention que l'expérience avec le substrat  $^{12}\text{C}$ . Cependant, le spectre de masse ne

correspond pas à celui du substrat où un pic de masse à 883 Da est attendu (Figure 76). Nous pouvons cependant émettre l'hypothèse que le substrat est bien synthétisé au vu du temps de rétention du seul pic observé, mais qu'il y a eu un problème au niveau de la mesure du spectre de masse. Une nouvelle analyse est à envisager afin de confirmer que le CoA-dikétide  $^{13}\text{C}$  a été correctement synthétisé.

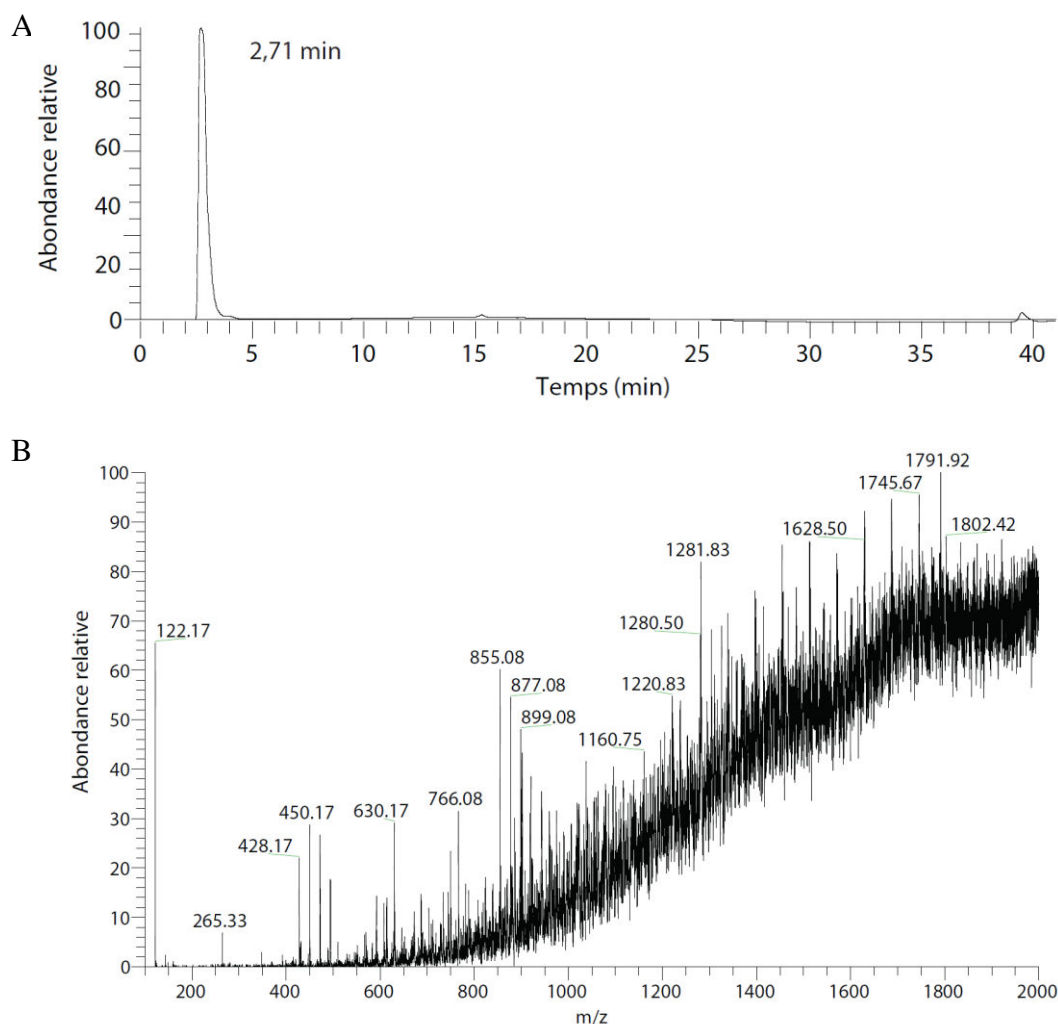


Figure 76 : Analyse par LC-MS du CoA-dikétide  $^{13}\text{C}$  synthétisé. A : partie chromatographie en phase liquide. B : spectrométrie de masse.

## IV.2 Modification de l'ACP<sub>1</sub> par le CoA-dikétide et analyse par spectrométrie de masse

### IV.2.1 Modification de l'ACP<sub>1</sub>

Le CoA-dikétide  $^{12}\text{C}$  a été utilisé pour modifier l'ACP<sub>1</sub>  $^{13}\text{C}$   $^{15}\text{N}$ , tandis que le CoA-dikétide  $^{13}\text{C}$  a été utilisé pour modifier l'ACP<sub>1</sub> non marqué. Le protocole a été le même que celui utilisé pour les modifications avec le CoA. La protéine et ses substrats ont été

placés en présence de Sfp, MgCl<sub>2</sub> et DTT. Les réactions se sont déroulées sur la nuit. Les deux types d'ACP<sub>1</sub> modifiés ont été par la suite purifiés par filtration sur gel.

#### IV.2.2 Vérification de la modification par spectrométrie de masse

Les analyses de spectrométrie de masse ont été à nouveau réalisées en collaboration avec C. Paris. Concernant l'ACP<sub>1</sub> modifié avec le CoA-dikétide <sup>13</sup>C, pour la partie chromatographie il n'y a qu'un seul pic avec un temps de rétention de 13,27 min. Pour la partie spectrométrie de masse, deux entités sont visibles. Les pics de masse à 1515,50 et 1767,92 m/z correspondent à une masse de 10601 Da (la masse de l'ACP<sub>1</sub> sous forme *apo*) à un état de charge +6 et +7. Les pics de masse à 1564,17 et 1824,67 m/z correspondent à une masse de 10941 Da (l'ACP sous forme *holo* mais non acylé) à un état de charge +6 et +7. Ces résultats suggèrent à la fois que la réaction de modification n'est pas totale et que lors de la modification, il y a eu lyse de l'acyle lié au bras phosphopantéthéine (Figure 77).

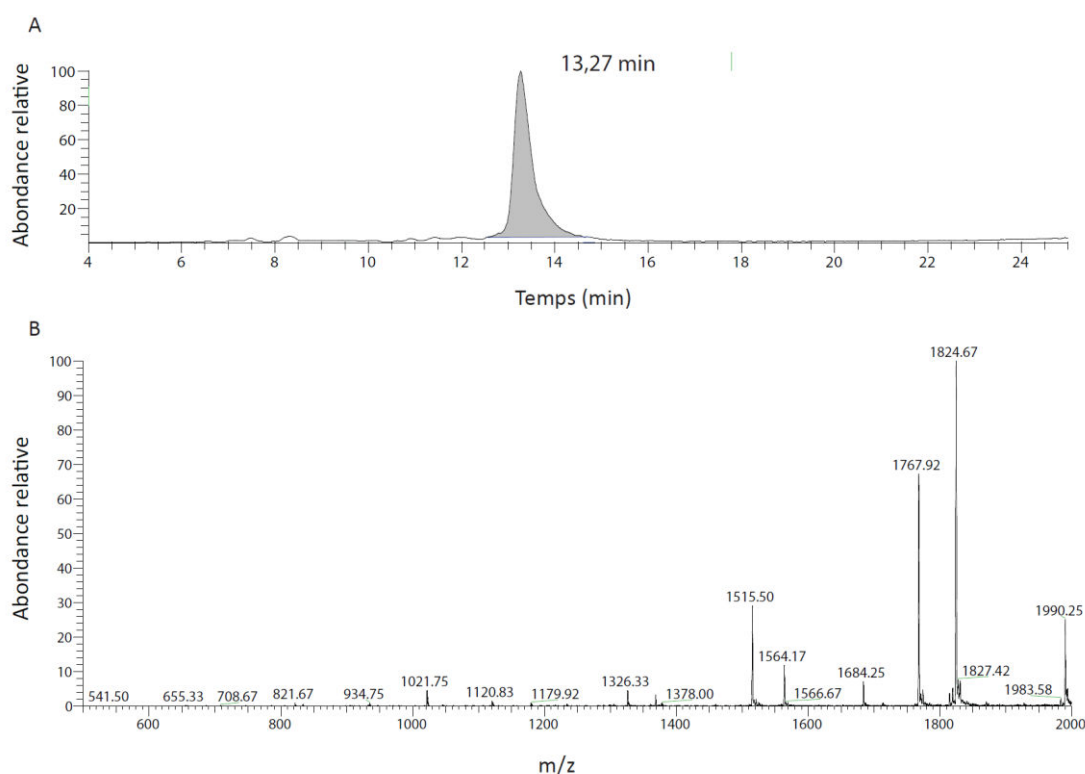


Figure 77 : Analyse par LC-MS de l'ACP<sub>1</sub> modifié avec le CoA-dikétide <sup>13</sup>C. A : partie chromatographie en phase liquide. B : spectrométrie de masse.

Concernant l'ACP<sub>1</sub> <sup>13</sup>C <sup>15</sup>N modifié avec le CoA-dikétide <sup>12</sup>C, pour la partie chromatographie il n'y a qu'un seul pic avec un temps de rétention de 13,21 min. Ce



temps de rétention est proche de celui de l'ACP<sub>1</sub> <sup>12</sup>C modifié avec le CoA-dikétide <sup>13</sup>C. Pour la partie spectrométrie de masse, il y a un pic de masse majoritaire à 1919,67 m/z correspondant à un état de charge +6 de la protéine marquée liée au bras phosphopantéthéine (Figure 78). Ces résultats suggèrent que la protéine est bien modifiée, mais que le dikétide est lysé.

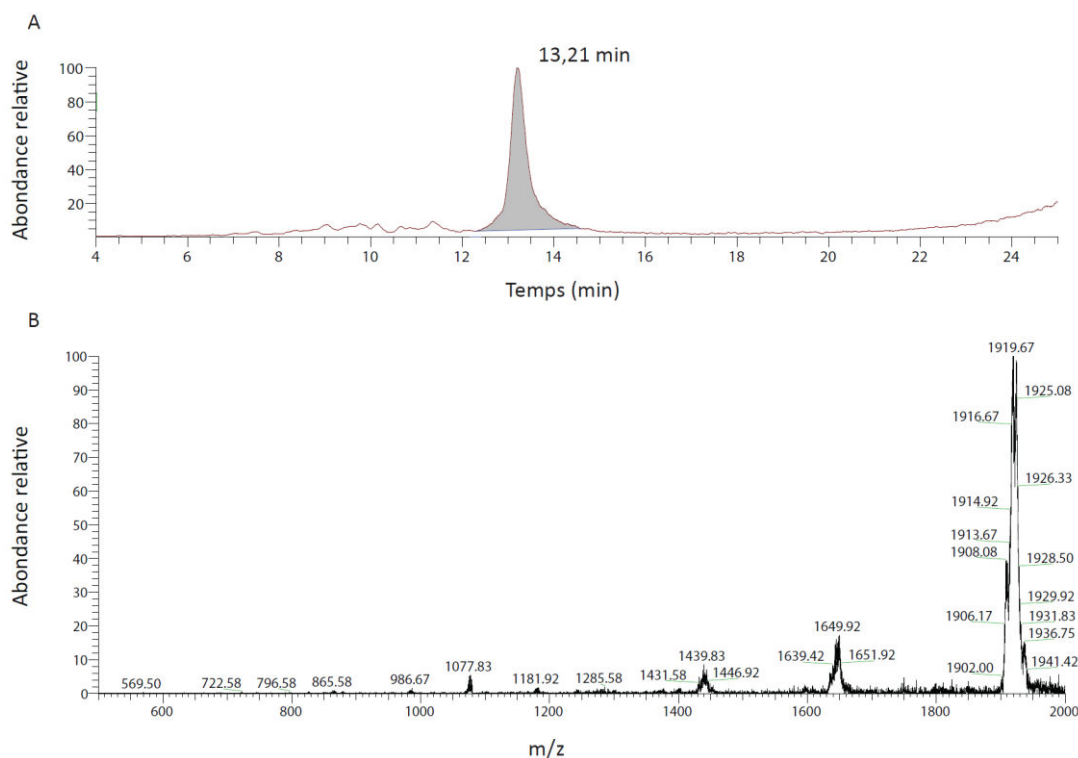


Figure 78 : Analyse par LC-MS de l'ACP<sub>1</sub> <sup>13</sup>C <sup>15</sup>N modifié avec le CoA-dikétide <sup>12</sup>C. A : partie chromatographie en phase liquide. B : spectrométrie de masse.

Deux problèmes ont donc été rencontrés durant la modification de l'ACP<sub>1</sub> avec les CoA-dikétides. Tout d'abord, il s'avère que les ACP<sub>1</sub> ont pu être modifiés, mais seulement avec le bras phosphopantéthéine, la partie dikétide a été lysée. La lyse pourrait être la conséquence de la présence de DTT dans le mélange réactionnel de la modification, qui aurait attaqué la fonction thiol du bras phosphopantéthéine, libérant ainsi le dikétide. Ensuite, le deuxième problème provient de l'analyse de l'ACP modifié avec le CoA-dikétide <sup>13</sup>C. En effet il y avait une population d'ACP non modifié et donc sous forme *apo*. Ceci pourrait être provoqué par une surestimation de la concentration en CoA-dikétide <sup>13</sup>C synthétisé.

Nous avons voulu, par la suite, contrôler l'intégrité des substrats pantéthéine-dikétide synthétiques (<sup>13</sup>C et <sup>12</sup>C) afin de nous assurer que les problèmes rencontrés avec le CoA-dikétide lysé ne proviendraient pas du produit de départ. La pantéthéine-

dikétide  $^{12}\text{C}$  a été analysé en premier et le chromatogramme montre deux pics avec des temps de rétention de 3,82 min et 5,47 min. Le chromatogramme suggère dans un premier temps qu'il y a deux populations de molécules. Un spectre de masse a été obtenu pour chacun des deux pics. Le spectre de masse de l'espèce à 3,82 min possède deux pics de masse principaux : un à 521 Da et l'autre à 261 Da. La masse à 261 Da correspond à la masse de la pantéthéine-dikétide auquel il manque l'acide pantoïque. La masse à 521 Da correspond à la forme dimérique de cette molécule. Le spectre de masse de l'espèce à 5,47 min possède également deux pics à 391,33 Da et 780,83 Da. Celui à 391,33 Da correspond à la masse de la pantéthéine-dikétide et celui à 780,83 Da à la forme dimérique du substrat. A noter qu'un pic à 373,50 Da est également présent. Cette masse est celle du substrat déshydraté (Figure 79).

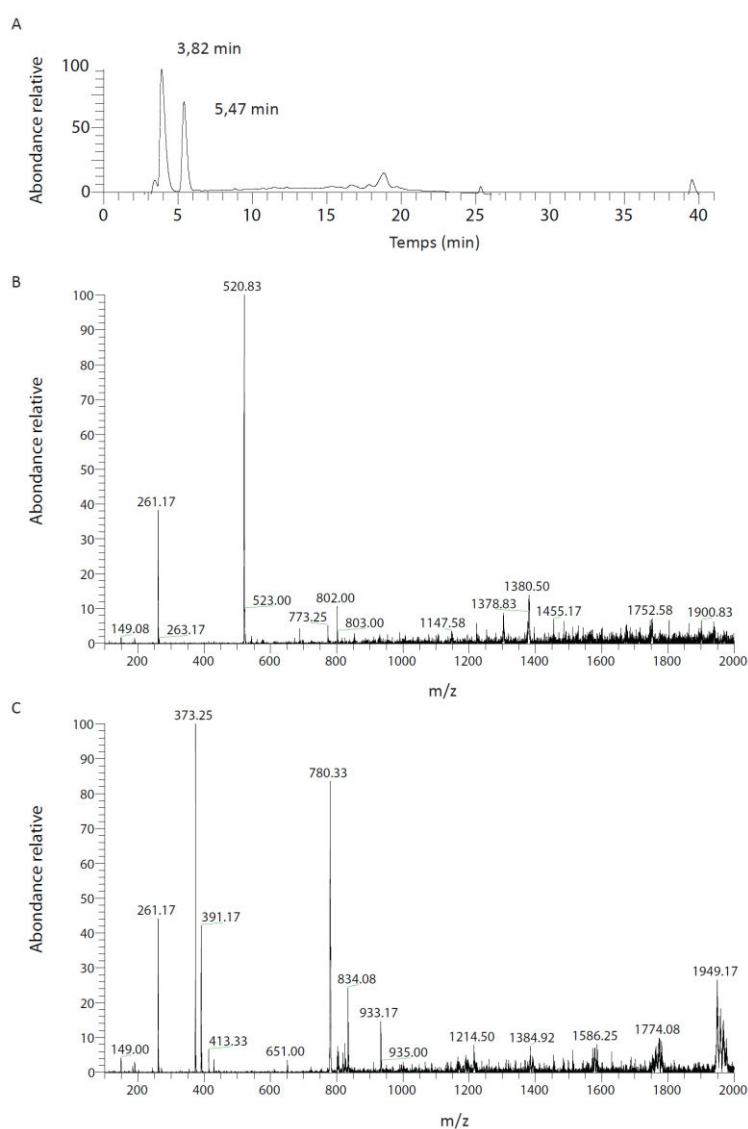


Figure 79 : Analyse par LC-MS de la pantéthéine-dikétide  $^{12}\text{C}$ . A : Chromatogramme. B : Spectre de masse du pic à 3,82 min. C : Spectre de masse du pic à 5,47 min.

Concernant la pantéthéine-dikétide  $^{13}\text{C}$ , son chromatogramme montre également deux pics à des temps de rétention de 3,86 min et 5,50 min, similaires à ceux du substrat non marqué. Le spectre de masse du premier pic d'élution montre deux pics principaux avec une masse de 264,17 Da et 526,92 Da. La première masse correspond à la masse de la pantéthéine-dikétide marqué auquel il manque l'acide pantoïque et la deuxième masse correspond au dimère de cette molécule lysée. Le spectre de masse du deuxième pic présente quant à lui, un pic majoritaire à une masse de 376,33 Da qui correspond à la masse de la pantéthéine-dikétide  $^{13}\text{C}$  (Figure 80).

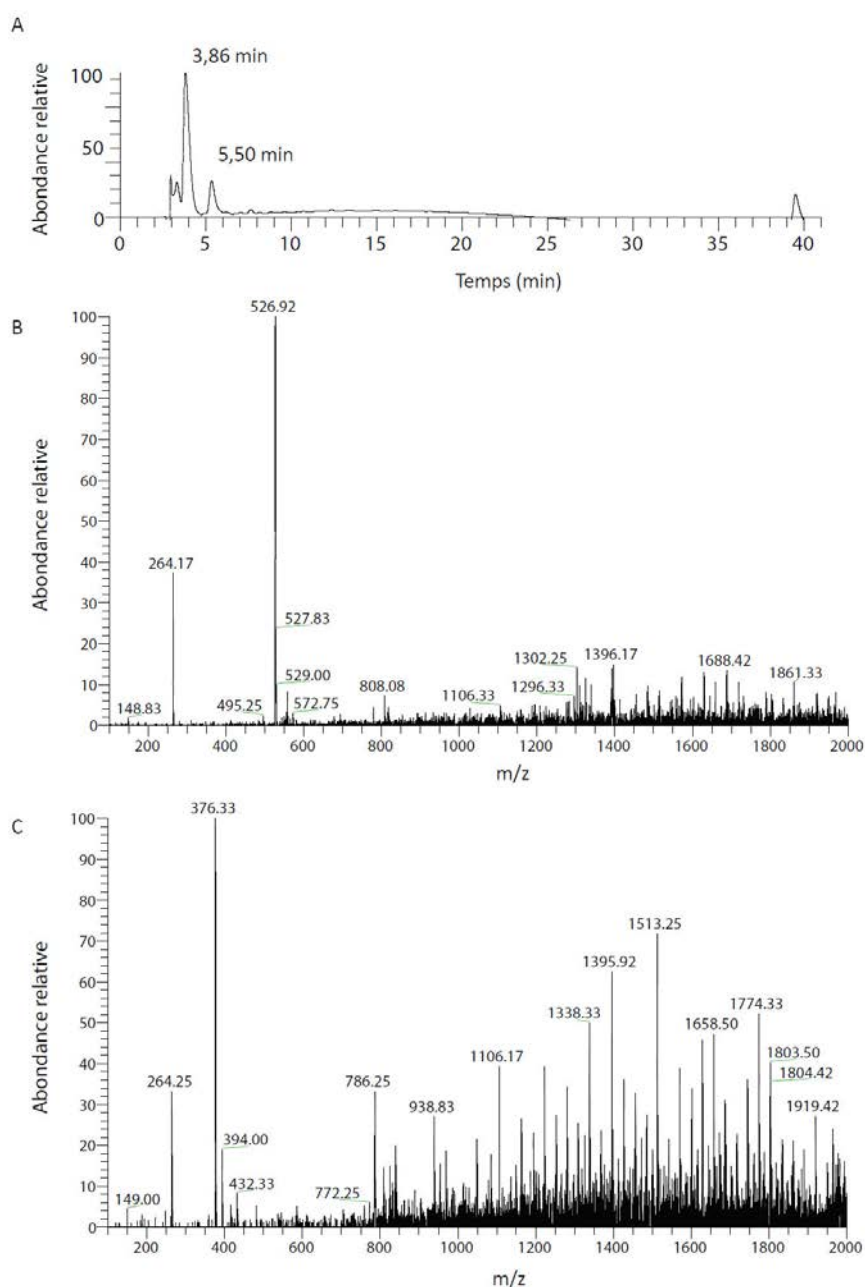


Figure 80 : Analyse LC-MS de la pantéthéine-dikétide  $^{13}\text{C}$ . A : Chromatogramme. B : Spectre de masse du pic à 3,86 min. C : Spectre de masse du pic à 5,50 min.

L'analyse par spectrométrie de masse des deux substrats pantéthéine-dikétides montre que ceux-ci sont dégradés. La proportion intacte de pantéthéine-dikétide  $^{12}\text{C}$  est de 28%, alors que celle de pantéthéine-dikétide  $^{13}\text{C}$  n'est que de 19%. Le reste est donc la forme avec l'acide pantoïque en moins (Figure 81). La pantéthéine-dikétide étant le seul substrat de l'enzyme PanK pour la première étape de synthèse du CoA-dikétide, le rendement de la réaction est réduit du fait de sa faible proportion. Par conséquent, il sera nécessaire d'augmenter la quantité de CoA-dikétide ajouté pendant la réaction de dérivatisation, afin de modifier l'ensemble des ACP<sub>1</sub>.

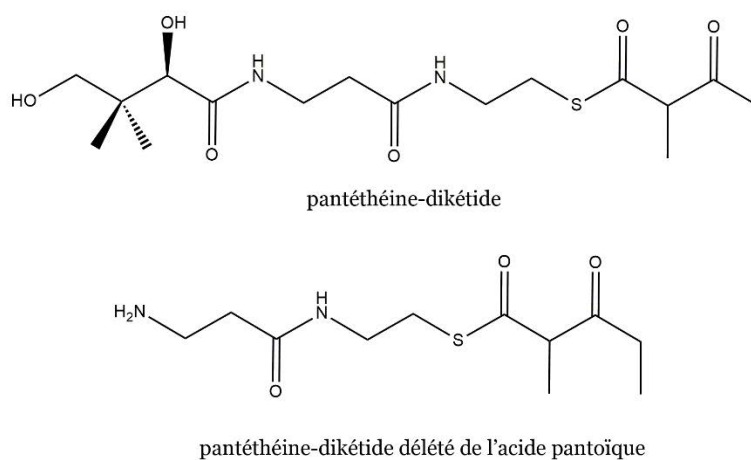


Figure 81 : Structure du substrat pantéthéine-dikétide et de sa forme tronquée de l'acide pantoïque.

*Discussion /  
perspectives*

## Discussion/perspectives

### Caractérisation d'une nouvelle classe de domaine de docking chez les PKS *trans*-AT

Les travaux de Davison *et al.* ont suggéré que les extrémités Cter de VirA et Nter de VirFG étaient des DD<sup>29</sup>. Sachant qu'à ce jour les DD caractérisés proviennent de PKS de type *cis*-AT, nous avons donc décidé d'obtenir la structure des DD d'une PKS *trans*-AT et de caractériser leur interaction. Dans un premier temps, nous avons voulu prouver que VirA<sup>CDD</sup> et VirFG<sup>NDD</sup> entraient bien en interaction. Cela a été confirmé par filtration sur gel. En effet, les volumes d'élution de VirA<sup>CDD</sup>, de VirFG<sup>NDD</sup> et des deux DD dans un mélange équimolaire sont tous les trois différents. Les poids moléculaires de chaque protéine ont été estimés avec le volume d'élution : celui de VirA<sup>CDD</sup> à 11,8 kDa (poids d'un monomère : 4,3 kDa), celui de VirFG<sup>NDD</sup> à 7,4 kDa (poids d'un monomère : 4,1 kDa) et le complexe VirA<sup>CDD</sup>/ VirFG<sup>NDD</sup> à 9,9 kDa (poids du complexe hétérodimérique : 8,4 kDa). Ces données suggèrent que lorsque VirA<sup>CDD</sup> est isolé il est sous forme oligomérique, alors que VirFG<sup>NDD</sup> isolé est monomérique même si son poids moléculaire est supérieur à celui d'un monomère. En effet, le poids moléculaire est surestimé à cause de l'élution précoce de VirFG<sup>NDD</sup> due à sa forme allongée. Pris ensemble, ils forment un complexe 1:1. L'affinité de l'interaction entre les deux partenaires a également été mesurée par ITC et le  $K_d$  obtenu a été de  $5,8 \pm 0,2 \mu\text{M}$ . Il se situe dans la gamme de  $K_d$  obtenue pour les DD de classe 1 et 2 (de 4 à 20  $\mu\text{M}$ )<sup>80,81,84</sup>. La mesure d'affinité a également été effectuée avec VirA<sup>ACP<sub>5b</sub>-CDD</sup> et VirFG<sup>NDD</sup> et également VirA<sup>CDD</sup> et VirFG<sup>NDD</sup> lié à la séquence linker en aval allant jusqu'au KS<sub>6</sub> (VirFG<sup>NDD</sup>-linker). Les  $K_d$  sont respectivement de 4 et 5,5  $\mu\text{M}$  et indiquent que l'ACP<sub>5b</sub> ainsi que le linker ne sont pas impliqués dans l'interaction.

Les DD isolés ont ensuite été caractérisés par RMN et par SAXS. Un spectre <sup>1</sup>H-<sup>15</sup>N HSQC de VirA<sup>CDD</sup> marqué <sup>15</sup>N a été enregistré. La faible proportion de pics visibles (15 pics ont été observés sur les 37 attendus) est due vraisemblablement à un équilibre entre dimères non-spécifiques de VirA<sup>CDD</sup>. Ceci est en accord avec les données SAXS qui montrent que ce domaine est homodimérique<sup>106</sup>. Le fait que VirA<sup>CDD</sup> se dimérise est confirmé par des études par ITC qui donnent également une estimation de la

concentration au-dessus de laquelle la protéine est complètement homodimérique (30  $\mu\text{M}$ ). En ajoutant VirFG<sup>NDD</sup> non marqué, VirA<sup>CDD</sup> forme un complexe hétérodimérique stable. En effet, le nombre de pics passe à 36 sur le spectre <sup>1</sup>H-<sup>15</sup>N HSQC enregistré.

A l'inverse, un spectre <sup>1</sup>H-<sup>15</sup>N HSQC de VirFG<sup>NDD</sup> marqué <sup>15</sup>N a été enregistré. Les pics attendus sont tous visibles mais ils présentent une faible dispersion de déplacements chimiques, suggérant que VirFG<sup>NDD</sup> n'est pas structuré. En ajoutant VirA<sup>CDD</sup> non marqué il y a des changements de déplacements chimiques et une dispersion globale des pics. Ces résultats laissent donc penser que VirFG<sup>NDD</sup> seul n'est pas structuré, alors qu'il l'est en présence de son partenaire. Ces données indiquent que VirFG<sup>NDD</sup> est une protéine intrinsèquement désordonnée. Cette observation est soutenue par des données de CD, où le pourcentage de structures secondaires pour la protéine seule augmente en présence de TFE (trifluoroéthanol). De plus, les données par SAXS montrent que VirFG<sup>NDD</sup> est une protéine partiellement structurée avec un  $D_{max}$  de 66 Å et que la protéine n'est pas globulaire au vu de l'allure du Kratky plot. A ce jour, cette protéine est un bon candidat pour être la première IDP participant au métabolisme secondaire bactérien. Cependant il aurait fallu confirmer (par CD par exemple) que le domaine KS n'induit pas le repliement du VirFG<sup>NDD</sup> en utilisant la protéine VirFG<sup>NDD</sup>-KS<sub>6</sub>. Néanmoins, jusqu'à présent, la protéine VirFG<sup>NDD</sup>-KS<sub>6</sub> s'est toujours agrégée lors des essais de purification.

Afin d'obtenir la structure des DD en interaction, nous avons fusionné les DD par un linker composé de deux répétitions GGGs. Ce linker est le même que celui utilisé lors de l'étude des DD de classe 2<sup>80</sup>. La structure, qui a été obtenue par RMN, a montré que chaque DD est composé de deux hélices  $\alpha$ . La structure globale des DD en interaction forme, quant à elle, un fagot de 4 hélices  $\alpha$ . Contrairement aux DD de classe 1 et 2 pour lesquels l'interaction se fait entre deux homodimères, l'interaction des sous-unités de ce système est composée d'une seule copie du DD Cter et Nter. Etant donné que les PKS sont dimériques, l'interface est formée de deux complexes de DD bien distincts. Cette classe de DD provenant d'un système *trans*-AT, est donc différente de celles identifiées jusqu'à présent au sein des systèmes *cis*-AT et représente notamment un nouveau type d'interaction protéine-protéine. La validité de la structure a été confirmée par SAXS. En effet, la courbe de diffusion générée par CRYSOLOG à partir de la structure RMN se superpose parfaitement à celle obtenue par SAXS.

La structure du complexe est maintenue grâce à des interactions hydrophobes au niveau de l'interface des deux DD, mais également entre les deux hélices  $\alpha$  de chaque DD. Trois interactions polaires sont également présentes à l'interface. Nous avons testé la pertinence de la structure des DD fusionnés en mesurant l'interaction entre les protéines discrètes. Pour cela, nous avons muté des résidus présents à l'interface des deux DD afin de quantifier leur impact sur l'affinité. Nous nous sommes concentrés sur des mutants VirA<sup>CDD</sup>, puisque VirFG<sup>NDD</sup> n'étant pas structuré seul, il est impossible de savoir si les mutants impactent son interaction avec VirA<sup>CDD</sup> ou son repliement induit. Une fois les résidus D6940, N6949 et E6954 mutés, l'affinité d'interaction mesurée par ITC est diminuée, ce qui implique que les interactions polaires sont importantes pour l'assemblage des DD. Ces résidus ne sont pas strictement conservés d'un couple de DD à l'autre chez un même système de PKS, ce qui permet une spécificité d'interaction et donc d'assurer l'ordonnement de la synthèse du polykétide. De plus, au niveau de ces DD, l'identité de séquence est plus faible que pour le reste de la sous-unité, ce qui suggère que c'est le caractère de l'acide aminé qui est primordial et non pas une séquence d'acides aminés spécifique.

Par la suite, nous avons voulu déterminer si les domaines ACP et KS situés de part et d'autres influent sur l'interaction, comme c'est le cas pour les *cis*-AT PKS<sup>81</sup>. L'analyse par SAXS de la fusion VirA ACP<sub>5b</sub>-<sup>CDD</sup>-VirFG<sup>NDD</sup> montre que l'ACP n'entre pas en contact avec les DD en interaction. Les calculs de forme *ab initio* obtenus par SAXS pour VirA ACP<sub>5b</sub>-<sup>CDD</sup> et VirFG<sup>NDD</sup>-KS<sub>6</sub> fusionnés montrent que le linker situé entre la KS<sub>6</sub> et le VirFG<sup>NDD</sup> est structuré. Ce linker pourrait être une IDP dans son contexte modulaire.

Nous avons caractérisé dans cette étude la première structure de DD d'un système *trans*-AT et définit ainsi une quatrième classe de DD. Celle-ci est composée d'hélices  $\alpha$  qui dévoilent une nouvelle topologie. Un des deux DD est probablement une IDP dont le repliement est induit par son partenaire. Il s'agit de la première IDP décrite chez les protéines responsables de la biosynthèse de métabolite secondaire bactérien. L'utilisation d'une IDP permet d'assurer une haute spécificité d'interaction tout en présentant une faible affinité. Ceci permet d'assurer l'ordonnement de la synthèse tout en facilitant le renouvellement d'une sous-unité défectueuse.



Grâce aux données SAXS obtenues sur la fusion VirA ACP<sub>5b</sub>-<sup>C</sup>DD–VirFG<sup>N</sup>DD-KS<sub>6</sub> nous avons pu reconstruire l'interface entre deux sous-unités de PKS *trans*-AT. En effet, l'ACP<sub>5b</sub> de la structure du module 5 se superpose parfaitement avec celui de la fusion<sup>29</sup>. Le module 5 possède une structure ouverte mais l'interaction avec la KS<sub>6</sub> suivante (grâce aux DD) ferme cette structure, afin de former une chambre de réaction. Cependant la distance entre l'ACP<sub>5b</sub> et la KS<sub>6</sub> de 80 Å est trop importante pour un transfert de chaîne entre les deux domaines. Un changement de conformation est donc essentiel à cette interaction. Celui-ci pourrait être induit par l'acylation du domaine ACP, comme au niveau du module 5 de la PKS Pik<sup>54</sup>. Ces changements seraient orchestrés grâce à la flexibilité des linkers entre les domaines. Des études futures devront consister à modifier l'ACP<sub>5b</sub> et de résoudre la structure du module 5 fusionné au module 6 par SAXS afin d'observer ces changements conformationnels.

Enfin, toujours dans une optique de biologie synthétique, l'utilisation de cette classe de DD pour créer une nouvelle interface nécessitera l'échange d'un couple de DD. Sachant qu'il n'est pas possible de prédire l'impact sur la structuration d'une IDP lorsqu'un résidu est muté, il est préférable d'échanger entièrement les couples de DD de cette classe plutôt qu'effectuer des mutations ponctuelles. De plus, d'après la structure SAXS, les ACP et KS n'interagissent pas, ce qui permettra en principe de créer de nouvelles interfaces par simple échange de domaine DD à condition que la spécificité de substrat du KS en aval soit respectée<sup>24,31</sup>.

### **L'échange de domaine KR comme moyen d'altérer la stéréochimie**

Nous avons voulu insérer une épimérisation au niveau d'un polykétide par simple échange de KR. Nous avons donc échangé, au sein de la PKS modèle DEBS 1-TE, la KR de type A1 du deuxième module par l'ensemble des KR épimérisantes caractérisées à ce jour et qui sont retrouvées au niveau de modules ne comprenant que ce domaine de réduction. Par cette technique, nous avons réussi à modifier la stéréochimie en introduisant une épimérisation.

Dans un premier temps, l'obtention de trikétide pour 14 constructions sur 16 suggère que le choix des sites de restriction utilisés pour insérer les séquences codant les KR a été judicieux. L'insertion des séquences codant les KR<sub>1</sub> en utilisant les sites de restriction *Bgl*III/*Nhe*I semble être une meilleure solution (avec des rendements

globaux plus importants) qu'avec le couple de sites de restriction *AvrII/NheI*. L'obtention de produits met également en avant que le domaine TE est capable de libérer un produit avec des changements de stéréochimie. De plus, les rendements globaux allant de traces à 59 mg/L sont semblables à ceux obtenus lors d'études antérieures qui allaient de 5,5 à 27 mg/L<sup>68</sup>. L'ensemble de ces faits met en avant que l'échange de domaines KR n'a pas altéré le bon fonctionnement de la PKS.

Des KR aux substrats très variés ont été insérées dans la PKS modèle. Cependant, aucune corrélation n'a pu être mise en place entre les rendements obtenus ou encore l'efficacité de réduction et la longueur de chaîne du substrat natif de chaque KR. En effet, les Amph KR 1 et 11 présentent les rendements les plus élevés en polykétide épimérisé. Cependant, alors que ces deux domaines possèdent comme substrat dans leur contexte natif un dikétide et un dodécakétide, respectivement, le substrat dans la PKS modèle est un trikétide. Puisque les domaines KR semblent présenter une large spécificité de substrat, leur interconversion reste possible dans le cadre d'études de biologie synthétique.

En échangeant 9 KR de type A2, nous avons pour 7 des PKS résultantes, l'obtention d'une lactone épimérisée. L'identification de la lactone épimérisée implique que l'échange de KR permet de modifier la stéréochimie des C $\alpha$ . Cependant, cette lactone est obtenue sous forme de traces pour 5 des PKS et à des concentrations de 7 et 14 mg/L lors de l'échange des deux KR Amph KR 1 et 11. Ainsi, l'échange de KR non-épimérisantes par des KR épimérisantes permet de modifier la stéréochimie des C $\alpha$  mais de manière peu efficace. La lactone non épimérisée, correspondant au produit natif de la PKS modèle, est également co-obtenue ainsi que la cétolactone qui est le produit non réduit, pour 7 de ces 9 KR échangées. Le fait d'obtenir un produit réduit mais non-épimérisé suggère que l'épimérisation n'est pas une condition absolue pour la réduction par ce type de KR. De plus, les KR ne génèrent normalement qu'un seul produit qui doit être épimérisé. L'obtention de plusieurs produits tend donc à confirmer que le module entier a un effet sur la stéréochimie des produits. La présence de la cétolactone signifie, quant à elle, que la PKS est fonctionnelle malgré l'échange de domaine, mais que la KR n'est pas toujours très efficace dans ce module de type A1.

En échangeant 7 KR de type B2, seulement 2 constructions ont donné un produit épimérisé. Cette faible proportion est vraisemblablement due au fait que les KR de ce

type doivent épimériser mais également modifier le sens de réduction par rapport à celui du module natif. De plus le produit attendu est uniquement obtenu sous forme de traces. Néanmoins l'ensemble des PKS a produit la cétolactone. Ces deux points montrent que les KR de type B2 sont peu efficaces dans ces types de module.

Dans le détail, les KR de type A2 montrent une portabilité accrue par rapport aux KR de type B2 dans ces modules de type A1. Cependant, plus de KR de type B2 devraient pouvoir réduire les polykétides puisque le sens de réduction est intrinsèque à la KR. La différence de portabilité entre les deux types de KR est due à la réaction d'épimérisation. En effet, cette réaction dépend non seulement de la KR mais aussi du contexte modulaire. En effet le contexte modulaire est essentiel pour l'épimérisation, dans le sens où la réaction peut être liée à l'accessibilité de la molécule d'eau, impliquée potentiellement dans le mécanisme, au site actif de la KR. De plus, des études de protéolyse limitée sur la PKS modèle ont mis en avant que le module 2 est plus résistant à la protéolyse que le module 1 (qui est un module épimérisant)<sup>107</sup>. Ces résultats tendent à montrer que le module 2 est plus compact que le module 1 (de type B2). Cette architecture et ce changement d'environnement modulaire peuvent apporter une difficulté supplémentaire afin de modifier la stéréochimie par changement de domaine.

De plus, hormis pour deux KR de type A2, l'épimérisation se fait de manière peu efficace au vu des rendements obtenus. Il en est de même avec la réaction de réduction caractérisée par la présence quasiment omniprésente de la cétolactone. De surcroît, à l'exception de trois KR, chaque fois que la cétolactone est synthétisée, il s'agit du produit majoritaire. Alors que, lors de l'insertion dans ce même module de plusieurs KR non épimérisantes, de type A1 et même B1, les produits de stéréochimie en lien avec les propriétés de la KR, avaient été obtenus majoritairement et ce avec des rendements élevés allant de 25 à 100 mg/L<sup>68</sup>.

Cette étude a montré que peu de KR sont capables d'épimériser efficacement lorsqu'elles sont introduites dans un contexte non-épimérisant. Pour généraliser ces données, une stratégie serait d'identifier et de tester de nouvelles KR épimérisantes grâce au séquençage et l'analyse des génomes. Bien que notre méthode permette d'arriver à un produit de stéréochimie souhaitée, d'autres produits sont synthétisés par les différentes PKS hybrides. Ces produits n'ont pas une stéréochimie corrélée avec les propriétés intrinsèques des domaines KR mais plutôt avec le contexte modulaire. Ainsi,

la stratégie future serait d'échanger des modules complets contenant des KR épimérisantes afin de s'affranchir des contraintes du contexte modulaire. Cette stratégie aura pour but de modifier plus efficacement la stéréochimie des carbones  $\alpha$  et  $\beta$  par rapport au simple changement de domaines KR.

### Essais de résolution de la structure d'une KR épimérisante

Il a été clairement identifié que les KR réalisent la réaction d'épimérisation, effectuée au niveau des  $C_\alpha$  en plus de leur activité de réduction<sup>46,60</sup>. Cependant les résidus réalisant cette réaction ne sont pas encore identifiés. Plusieurs structures de domaines KR ont été résolues, mais elles n'ont pas permis d'identifier le résidu (ou les résidus) catalysant l'épimérisation, contrairement à ceux intervenant dans la réduction. Nous avons donc décidé de tenter de résoudre la structure d'une KR épimérisante d'un premier module d'une PKS en présence de son substrat natif dikétide lié à la pantéthéine/phosphopantéthéine, ou d'un analogue (la *trans*-1-décalone), et du cofacteur réduit NADPH<sub>3</sub>, afin d'obtenir un complexe ternaire. Auparavant, notre équipe s'était penchée sur la DEBS KR<sub>1</sub> (H. Le Cordier, données non-publiées). En effet, la structure de cette KR a été résolue en présence de cofacteur mais pas de substrat<sup>43</sup>. Ainsi, notre équipe a tenté de reproduire les conditions de cristallisation publiées pour ensuite réaliser le trempage du cristal en présence du substrat et du cofacteur réduit. Malheureusement, la protéine n'a pas pu être cristallisée. Nous nous sommes donc intéressés à d'autres KR : les KR<sub>1</sub> des PKS pikromycine et lankamycine. En plus des KR<sub>1</sub>, nous avons voulu résoudre la structure de la KR<sub>A</sub> de la PKS mycolactone, qui possède une large spécificité de substrat<sup>108</sup> mais également la KR<sub>6</sub> de la PKS virginiamycine afin d'obtenir la structure d'une KR d'une PKS *trans*-AT.

Nous avons ainsi tenté de cristalliser ces quatre KR grâce aux différents kits de cristallisation. Nous n'avons obtenu des cristaux que pour la KR<sub>1</sub> de la PKS lankamycine en présence de TCEP et en absence de substrat et de NADPH<sub>3</sub>. Nous avons reproduit les cristaux pour cette KR, mais ceux-ci étaient toujours maclés. Pour pallier ce problème, des cristaux ont été produits par micro-ensemencement. Ainsi les cristaux obtenus sont au préalable trempés ou non dans une solution contenant les substrats et le cofacteur inactif, puisque la co-cristallisation a inhibé la formation de

cristaux, puis ils sont congelés. Afin d'avoir des cristaux pour les autres protéines, nous avons réalisé des tests en faisant varier la nature et la concentration de l'agent précipitant, la nature et le pH du tampon, et la nature des sels. Malgré cela, nous n'avons pas pu avoir de cristaux pour ces trois protéines.

La résolution de la structure a donc échoué. En effet, la faible proportion de solvant dans les cristaux est un frein pour phaser par remplacement moléculaire, d'autant que la qualité de l'empilement cristallin est médiocre. Nous avons alors décidé d'utiliser le signal anomal des lanthanides. Aucun atome n'a diffusé dans les cristaux par trempage et aucun dérivé n'a été obtenu par co-cristallisation malgré les fortes concentrations utilisées (jusqu'à 100mM). Ce résultat peut être expliqué par le faible pourcentage de solvant à l'intérieur des cristaux (40 %). De nombreux essais de cristallisation ont été menés pour avoir une autre forme cristalline, sans succès. Donc pour résoudre le problème de la phase avec ces cristaux, deux voies sont encore possibles. La première consisterait à introduire des méthionines par mutagenèse dirigée et ensuite de cristalliser la protéine sélénométhionylée, tout en espérant que la protéine ainsi modifiée cristallise toujours. La seconde, plus difficile, consisterait à utiliser les dommages d'irradiation induits par l'exposition aux rayons X<sup>109</sup>.

### **Etude de la stabilisation de la stéréochimie du C $\alpha$**

L'objectif de ce projet était d'identifier le domaine, au sein d'un module modèle, responsable de la stabilisation de la configuration du C $\alpha$  entre les réactions de condensation et de réduction.

Dans un premier temps, nous avons attribué la chaîne principale de l'ACP<sub>1</sub> de la PKS Pikromycine afin d'identifier les résidus dont l'environnement chimique est potentiellement perturbé par l'ajout du substrat. L'ensemble des NH a été attribué, à l'exception de 4 résidus invisibles dont la sérine qui porte le bras phosphopantéthéine. La sérine étant invisible, nous ne pouvions pas observer l'ajout du bras directement. Cependant l'environnement chimique d'autres résidus était perturbé par la présence du bras.

Ensuite, nous devons nous assurer que l'ACP<sub>1</sub> pouvait être modifiée. Pour cela, nous avons utilisé deux PPTases, la Sfp et la Svp, et le CoA seul comme substrat. Nous

avons confirmé par spectrométrie de masse que l'ACP<sub>1</sub> a été modifiée par les deux enzymes et que l'ajout du bras a été réalisé sur la bonne sérine. Par RMN, nous avons comparé les spectres de l'ACP<sub>1</sub> sous forme *apo* et *holo* et identifié les résidus dont l'environnement est perturbé.

Nous avons ensuite synthétisé les CoA-dikétides marqués <sup>13</sup>C ou non. L'ACP<sub>1</sub> marqué <sup>15</sup>N ou non a donc été modifié, et les spectres <sup>1</sup>H-<sup>15</sup>N HSQC et COCA ont ensuite été enregistrés. Aucune variation au niveau des spectres <sup>1</sup>H-<sup>15</sup>N HSQC de l'ACP<sub>1</sub> modifié avec le CoA ou le CoA-dikétide n'est visible. De plus, les pics des trois carbones <sup>13</sup>C ne sont pas visibles sur le spectre COCA. Nous avons tout d'abord vérifié par spectrométrie de masse que les substrats sont intègres. Par contre, en présence de CoA-dikétide, l'ACP<sub>1</sub> est soit modifiée par le CoA sans dikétide ou soit retrouvée sous forme *apo*. Concernant l'ACP<sub>1</sub> modifiée uniquement par le CoA, il y a eu lyse du dikétide. Sachant que lors de leur analyse, les deux CoA-dikétides étaient intègres (pas de CoA ou de dikétide seul identifié), la lyse s'est probablement produite lors de l'étape de modification. Nous devons recommencer la réaction de modification en absence de DTT qui peut être la cause de la lyse. Concernant l'ACP<sub>1</sub> retrouvée sous forme *apo*, il pourrait découler d'une concentration en substrat plus faible que celle calculée initialement. Nous avons décidé donc dans un premier temps d'analyser les produits de départ : les pantéthéine-dikétides <sup>12</sup>C ou <sup>13</sup>C. L'analyse par spectrométrie de masse a révélé que les deux substrats étaient dégradés à hauteur de 80 %. Ainsi lors de la réaction de modification, seulement 20 % de la concentration en substrat était du CoA-dikétide. Lors des modifications futures, afin d'obtenir quasiment 100 % d'ACP<sub>1</sub> modifiée, nous devons augmenter la concentration en substrat ajouté. Une autre solution sera de reconstituer les pantéthéine-dikétides dégradés. Pour ce faire, la protéine PanC (fournit par le Pr C. Abell, Cambridge) intervenant dans la voie de biosynthèse du CoA, devrait permettre de reconstituer la liaison lysée de la pantéthéine-dikétide.

En respectant ces conditions, nous devrions pouvoir réaliser la modification de l'ACP<sub>1</sub> de manière efficace et ainsi pouvoir apporter la confirmation que le domaine ACP ne stabilise pas la stéréochimie du C<sub>α</sub>. Pour arriver à cette conclusion nous devons comparer les spectres <sup>1</sup>H-<sup>15</sup>N HSQC du Pik ACP<sub>1</sub> <sup>15</sup>N *apo* et Pik ACP<sub>1</sub> <sup>15</sup>N *holo* acylé avec le dikétide afin d'observer si des déplacements de pics ont lieu. Nous devons également comparer les spectres 1D carbone et COCA du CoA-dikétide <sup>13</sup>C seul et lié à l'ACP<sub>1</sub> afin

d'observer des déplacements de pics éventuels. Par la suite, il sera intéressant de déterminer s'il y a des interactions entre le di-domaine KS<sub>1</sub>-AT<sub>1</sub> de Pik et le C<sub>α</sub> du substrat. Pour cela, nous modifierons dans un premier temps l'ACP<sub>1</sub> avec le CoA-dikétide <sup>13</sup>C. La protéine modifiée sera placée dans du D<sub>2</sub>O et des spectres 1D <sup>13</sup>C-édité <sup>1</sup>H seront enregistrés afin de mesurer la vitesse d'échange entre le deutérium et l'hydrogène lié au C<sub>α</sub>. Cette vitesse sera mesurée par la perte du signal de l'hydrogène. La même expérience pourra être réalisée en ajoutant un excès de di-domaine KS<sub>1</sub>-AT<sub>1</sub>. Si la vitesse d'échange est identique, alors le di-domaine KS<sub>1</sub>-AT<sub>1</sub> n'interagit pas avec le substrat, alors que si elle est moindre, cela suggèrera que le di-domaine KS<sub>1</sub>-AT<sub>1</sub> est en contact avec le substrat et qu'il stabilise sa stéréochimie. Cette même expérience sera réalisée en utilisant le CoA-dikétide deutéré en C<sub>α</sub> et en mesurant alors la vitesse d'échange *via* l'apparition d'un pic correspondant à l'hydrogène lié au C<sub>α</sub>.

Le même type d'expérience sera réalisé avec l'ACP<sub>1</sub> CoA-dikétide deutéré en C<sub>α</sub>, le di-domaine KS<sub>1</sub>-AT<sub>1</sub> et la Pik KR<sub>1</sub>. Les vitesses d'échange seront mesurées *via* l'apparition ou la disparition du pic correspondant à l'hydrogène en C<sub>α</sub> en absence et présence de NADPH<sub>3</sub> afin d'identifier si le domaine KR est l'agent stabilisateur de la stéréochimie.

*Références  
bibliographiques*



**Références bibliographiques**

1. Ghasemzadeh, A. & Jaafar, H. Effect of CO<sub>2</sub> Enrichment on Synthesis of Some Primary and Secondary Metabolites in Ginger (*Zingiber officinale* Roscoe). *Int. J. Mol. Sci.* **12**, 1101–1114 (2011).
2. Agostini-costa, T. S., Vieira, R. F., Bizzo, H. R., Silveira, D. & Gimenes, M. A. Secondary Metabolites. *Chromatogr. Its Appl.* **1**, 131–164 (2012).
3. Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* **41**, 185–201 (2014).
4. Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829–837 (2004).
5. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 1–12 (2015).
6. Wright, L. F. & Hopwood, D. A. Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **96**, 289–297 (1976).
7. Rudd, B. A & Hopwood, D. A. Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **114**, 35–43 (1979).
8. Craney, A., Ahmed, S. & Nodwell, J. Towards a new science of secondary metabolism. *J. Antibiot. (Tokyo)*. **66**, 387–400 (2013).
9. Weber, J. M., Leung, J. O., Maine, G. T., Potenz, R. H. B., Paulus, T. J. & DeWitt, J. P. Organization of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. *J. Bacteriol.* **172**, 2372–2383 (1990).
10. Cortes, J., Haydock, S. F., Roberts, G. A, Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178 (1990).
11. Weber, J. M., Schoner, B. & Losick, R. Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *Gene* **75**, 235–241 (1989).
12. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3**, 925–936 (2005).
13. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).

14. Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110 (2003).
15. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).
16. Hitchman, T. S., Crosby, J., Byrom, K. J., Cox, R. J. & Simpson, T. J. Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. *Chem. Biol.* **5**, 35–47 (1998).
17. Metz JG, Roessler P, Facciotti D, Levering C, Dittrich F, Lassner M, Valentine R, Lardizabal K, Domergue F, Yamada A, Yazawa K, Knauf V, B. J. Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science.* **293**, 290–293 (2001).
18. Van Lanen, S. G. & Shen, B. Biosynthesis of enediyne antitumor antibiotics. *Curr. Top. Med. Chem.* **8**, 448–459 (2008).
19. Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5**, 2010–2026 (2007).
20. Stinear, T. P., Mve-Obiang, A., Small, P. L. C., Frigui, W., Pryor, M. J., Brosch, R., Jenkin, G. A., Johnson, P. D. R., Davies, J. K., Lee, R. E., Adusumilli, S., Garnier, T., Haydock, S. F., Leadlay, P. F. & Cole, S. T. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1345–1349 (2004).
21. Yuzawa, S., Kapur, S., Cane, D. & Khosla, C. Role of a Conserved Arginine Residue in Linkers between the Ketosynthase and Acyltransferase Domains of Multimodular Polyketide Synthases. *Biochemistry* **51**, 3708–10 (2012).
22. Donadio, S., McAlpine, J. B., Sheldon, P. J., Jackson, M. & Katz, L. An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7119–7123 (1993).
23. Scotti, C., Piatti, M., Cuzzoni, A., Perani, P., Tognoni, A., Grandi, G., Galizzi, A. & Albertini, a. M. A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* **130**, 65–71 (1993).
24. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14002–14007 (2002).
25. Cheng, Y.-Q., Tang, G.-L. & Shen, B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–3154 (2003).
26. Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2010).

27. Johnston, N. J., Mukhtar, T. A & Wright, G. D. Streptogramin antibiotics: mode of action and resistance. *Curr. Drug Targets* **3**, 335–344 (2002).
28. Pulsawat, N., Kitani, S. & Nihira, T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*. *Gene* **393**, 31–42 (2007).
29. Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A. & Weissman, K. J. Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem. Sci.* **5**, 3081 (2014).
30. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207–220 (2001).
31. Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C. & Piel, J. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
32. Jenke-Kodama, H., Börner, T. & Dittmann, E. Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLoS Comput. Biol.* **2**, 1210–1218 (2006).
33. McDaniel, R., Thamchaipenet, A, Gustafsson, C., Fu, H., Betlach, M. & Ashley, G. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel ‘unnatural’ natural products. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1846–1851 (1999).
34. Kushnir, S., Sundermann, U., Yahiaoui, S., Brockmeyer, A., Janning, P. & Schulz, F. Minimally Invasive Mutagenesis Gives Rise to a Biosynthetic Polyketide Library. *Angew. Chemie Int. Ed.* **51**, 10664–10669 (2012).
35. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
36. Rowe, C. J., Böhm, I. U., Thomas, I. P., Wilkinson, B., Rudd, B. A M., Foster, G., Blackaby, A. P., Sidebottom, P. J., Roddis, Y., Buss, A. D., Staunton, J. & Leadlay, P. F. Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem. Biol.* **8**, 475–485 (2001).
37. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–11129 (2006).

38. Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., Challis, G. L. & Aigle, B. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci.* **108**, 6258–6263 (2011).
39. Tsai S., S. C. & Ames, B. D. *Chapter 2 Structural Enzymology of Polyketide Synthases. Methods in Enzymology* **459**, 17–47, (2009).
40. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931–943 (2007).
41. Witkowski, A., Joshi, A. K. & Smith, S. Mechanism of the beta-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. *Biochemistry* **41**, 10877–10887 (2002).
42. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–73 (2012).
43. Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737–48 (2006).
44. Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D. V., Richard Hutchinson, C. & McDaniel, R. A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79 (2003).
45. McPherson, M., Khosla, C. & Cane, D. E. Erythromycin Biosynthesis: The  $\beta$ -Ketoreductase Domains Catalyze the Stereospecific Transfer of the 4- pro - S Hydride of NADPH. *J. Am. Chem. Soc.* **120**, 3267–3268 (1998).
46. Garg, A., Xie, X., Keatinge-Clay, A., Khosla, C. & Cane, D. E. Elucidation of the cryptic epimerase activity of redox-inactive ketoreductase domains from modular polyketide synthases by tandem equilibrium isotope exchange. *J. Am. Chem. Soc.* **136**, 10190–3 (2014).
47. Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941–53 (2008).
48. Labonte, J. W. & Townsend, C. A. Active site comparisons and catalytic mechanisms of the hot dog superfamily. *Chem. Rev.* **113**, 2182–204 (2013).
49. Kwan, D. H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-Stark, J. C. C., Haydock, S. F. & Leadlay, P. F. Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **15**, 1231–1240 (2008).
50. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615–21 (2012).

51. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315–1322 (2008).
52. Tsai, S. C., Miercke, L. J., Krucinski, J., Gokhale, R., Chen, J. C., Foster, P. G., Cane, D. E., Khosla, C. & Stroud, R. M. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14808–14813 (2001).
53. Weissman, K. J. & Müller, R. A brief tour of myxobacterial secondary metabolism. *Bioorganic Med. Chem.* **17**, 2121–2136 (2009).
54. Dutta, S., Whicher, J. R., Hansen, D. A, Hale, W. A, Chemler, J. A, Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L. & Skinnotis, G. Structure of a modular polyketide synthase. *Nature* **510**, 512–7 (2014).
55. Whicher, J. R., Dutta, S., Hansen, D. A, Hale, W. A, Chemler, J. A, Dosey, A. M., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L. & Skinnotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–4 (2014).
56. Schlünzen, F., Zarivach, R., Harms, J., Bashan, A, Tocilj, A, Albrecht, R., Yonath, A & Franceschi, F. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
57. Marsden, A F., Caffrey, P., Aparicio, J. F., Loughran, M. S., Staunton, J. & Leadlay, P. F. Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **263**, 378–80 (1994).
58. Weissman, K. J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J. & Leadlay, P. F. The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. *Biochemistry* **36**, 13849–55 (1997).
59. Holzbaur, I. E., Ranganathan, A., Thomas, I. P., Kearney, D. J. A., Reather, J. A., Rudd, B. A. M., Staunton, J. & Leadlay, P. F. Molecular basis of Celmer's rules : role of the ketosynthase domain in epimerisation and demonstration that ketoreductase domains can have altered product specificity with unnatural substrates. *Chem. Biol.* **8**, 329–340 (2001).
60. Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **14**, 898–908 (2007).
61. Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spittler, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J. & Leadlay, P. F. Molecular basis of Celmer's rules: stereochemistry of catalysis by isolated ketoreductase domains from modular polyketide synthases. *Chem. Biol.* **12**, 1145–53 (2005).

62. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* **4**, 654–657 (2003).
63. Baerga-Ortiz, A., Popovic, B., Siskos, A. P., O'Hare, H. M., Spiteller, D., Williams, M. G., Campillo, N., Spencer, J. B. & Leadlay, P. F. Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase. *Chem. Biol.* **13**, 277–285 (2006).
64. Kwan, D. H., Tosin, M., Schläger, N., Schulz, F. & Leadlay, P. F. Insights into the stereospecificity of ketoreduction in a modular polyketide synthase. *Org. Biomol. Chem.* **9**, 2053–2056 (2011).
65. Bonnett, S. A., Whicher, J. R., Papireddy, K., Florova, G., Smith, J. L. & Reynolds, K. A. Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* **20**, 772–783 (2013).
66. Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C. & Cane, D. E. The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc.* **131**, 18501–18511 (2009).
67. Zheng, J. & Keatinge-Clay, A. T. Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases. *J. Mol. Biol.* **410**, 105–17 (2011).
68. Kellenberger, L., Galloway, I. S., Sauter, G., Böhm, G., Hanefeld, U., Cortés, J., Staunton, J. & Leadlay, P. F. A polylinker approach to reductive loop swaps in modular polyketide synthases. *Chembiochem* **9**, 2740–2749 (2008).
69. Castonguay, R., He, W., Chen, A. Y., Khosla, C. & Cane, D. E. Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **129**, 13758–69 (2007).
70. Evans, S. E., Williams, C., Arthur, C. J., Płoskoń, E., Wattana-amorn, P., Cox, R. J., Crosby, J., Willis, C. L., Simpson, T. J. & Crump, M. P. Probing the Interactions of Early Polyketide Intermediates with the Actinorhodin ACP from *S. coelicolor* A3(2). *J. Mol. Biol.* **389**, 511–528 (2009).
71. Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A. & Weissman, K. J. Insights into Protein-Protein and Enzyme-Substrate Interactions in Modular Polyketide Synthases. *Chem. Biol.* **17**, 705–716 (2010).
72. Charkoudian, L. K., Liu, C. W., Capone, S., Kapur, S., Cane, D. E., Togni, A., Seebach, D. & Khosla, C. Probing the interactions of an acyl carrier protein domain from the 6-deoxyerythronolide B synthase. *Protein Sci.* **20**, 1244–1255 (2011).
73. Maier, T., Jenni, S. & Ban, N. Architecture Synthase Acid of Mammalian Fatty A Resolution. **311**, 1258–1262 (2006).

74. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A, Murray, M., Hopwood, D. A & Simpson, T. J. Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **36**, 6000–6008 (1997).
75. Wu, N., Cane, D. E. & Khosla, C. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41**, 5056–66 (2002).
76. Tang, L., Fu, H. & McDaniel, R. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chem. Biol.* **7**, 77–84 (2000).
77. Tsuji, S. Y., Cane, D. E. & Khosla, C. Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **40**, 2326–2331 (2001).
78. Thattai, M., Burak, Y. & Shraiman, B. I. The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **3**, 1827–1835 (2007).
79. Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2008).
80. Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H. & Smith, J. L. Cyanobacterial polyketide synthase docking domains: A tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
81. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
82. Weissman, K. J. The structural basis for docking in modular polyketide biosynthesis. *ChemBioChem* **7**, 485–494 (2006).
83. Weissman, K. J. Single amino acid substitutions alter the efficiency of docking in modular polyketide biosynthesis. *ChemBioChem* **7**, 1334–1342 (2006).
84. Buchholz, T. J., Geders, T. W., Iii, F. E. B., Reynolds, K. A, Janet, L. & Sherman, D. H. Structural Basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem Biol* **4**, 41–52 (2009).
85. Chen, Y., Deng, W., Wu, J., Qian, J., Chu, J., Zhuang, Y., Zhang, S. & Liu, W. Genetic modulation of the overexpression of tailoring genes eryK and eryG leading to the improvement of erythromycin A purity and production in *Saccharopolyspora erythraea* fermentation. *Appl. Environ. Microbiol.* **74**, 1820–1828 (2008).

86. Yuan, Y., Chung, H. S., Leimkuhler, C., Walsh, C. T. & Walker, S. In vitro reconstitution of EryCIII activity for the preparation of unnatural macrolides. *J Am Chem Soc* **127**, 14128–14129 (2005).
87. Olano, C., Méndez, C. & Salas, J. A. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat. Prod. Rep.* **27**, 571–616 (2010).
88. Jiang, M., Zhang, H., Park, S. H., Li, Y. & Pfeifer, B. A. Deoxysugar pathway interchange for erythromycin analogues heterologously produced through *Escherichia coli*. *Metab. Eng.* **20**, 92–100 (2013).
89. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
90. Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal. Biochem.* **324**, 100–105 (2004).
91. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
92. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
93. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
94. Vagin, A. & Teplyakov, A. Molecular replacement with *MOLREP*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 22–25 (2010).
95. Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. *BALBES* : a molecular-replacement pipeline. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 125–132 (2008).
96. Girard, É., Stelter, M., Vicat, J. & Kahn, R. A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59**, 1914–1922 (2003).
97. Pompidor, G., D'Aléo, A., Vicat, J., Toupet, L., Giraud, N., Kahn, R. & Maury, O. Protein Crystallography through Supramolecular Interactions between a Lanthanide Complex and Arginine. *Angew. Chemie Int. Ed.* **47**, 3388–3391 (2008).
98. Cavanagh, J., Fairbrother, W., Palmer III, A., Rance, M. & Skelton, N. *Protein NMR spectroscopy. Principles and Practice*. Elsevier Academic Press **2nd**, (1996).



99. Annaïval, T., Paris, C., Leadlay, P. F., Jacob, C. & Weissman, K. J. Evaluating Ketoreductase Exchanges as a Means of Rationally Altering Polyketide Stereochemistry. *ChemBiochem* **16**, 1357–64 (2015).
100. Zheng, J., Fage, C. D., Demeler, B., Hoffman, D. W. & Keatinge-Clay, A. T. The missing linker: A dimerization motif located within polyketide synthase modules. *ACS Chem. Biol.* **8**, 1263–1270 (2013).
101. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
102. Zheng, J., Taylor, C. A., Piasecki, S. K. & Keatinge-Clay, A. T. Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **18**, 913–922 (2010).
103. Zheng, J. & Keatinge-Clay, A. T. Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases. *J. Mol. Biol.* **410**, 105–117 (2011).
104. Wong, H. C. The Solution Structure of Acyl Carrier Protein from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **277**, 15874–15880 (2002).
105. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
106. Fischer, H., De Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).
107. Hong, H., Appleyard, A. N., Siskos, A. P., Garcia-Bernardo, J., Staunton, J. & Leadlay, P. F. Chain initiation on type I modular polyketide synthases revealed by limited proteolysis and ion-trap mass spectrometry. *FEBS J.* **272**, 2373–2387 (2005).
108. Bali, S. & Weissman, K. J. Ketoreduction in mycolactone biosynthesis: Insight into substrate specificity and stereocontrol from studies of discrete ketoreductase domains in vitro. *ChemBioChem* **7**, 1935–1942 (2006).
109. Rice, L. M., Earnest, T. N. & Brunger, A. T. Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 1413–1420 (2000).

Les polykétides sont des composés naturels synthétisés par les bactéries, champignons et plantes. Ces composés possèdent des rôles thérapeutiques variés tels que antifongiques, antibiotiques, anticancéreux, immunosuppresseurs ou encore anticholestérolémiques. Ils représentent donc un important intérêt médical mais également économique. Par conséquent, la recherche de nouvelles structures possédant des bioactivités diverses se révèle être intéressante. Une stratégie prometteuse pour créer des nouveaux polykétides est l'ingénierie génétique des enzymes synthétisant ces molécules, les polykétide synthases modulaires (PKS), une approche désignée sous le terme de « biologie synthétique ». Pourtant, pour que cette approche soit efficace, il faut comprendre de façon détaillée le fonctionnement de ces systèmes multienzymatiques. Plusieurs points restent à éclaircir, dont : i) les interactions entre le domaine *acyl carrier protein* (ACP) porteur de la chaîne acyle en cours de synthèse et les domaines catalytiques, ainsi que la manière avec laquelle les domaines agissent ensemble au sein d'un module pour contrôler la stéréochimie du polykétide ; et ii) la façon dont les sous-unités présentes au sein de chaque PKS arrivent à communiquer entre elles de façon efficace et spécifique, ce qui implique l'étude des « domaines de docking » essentiels à ces interactions.

Lors de ma thèse, j'ai identifié deux kétoréductases (KR) qui, introduites dans un contexte modulaire intrinsèquement non-épimérisant, sont capables d'épimériser le méthyle en  $C_{\alpha}$  de façon efficace. J'ai mis en évidence que la modification de la stéréochimie du polykétide ne dépend pas exclusivement des propriétés intrinsèques de la kétoréductase (KR) mais qu'il faut prendre en compte le contexte modulaire dans lequel la KR se situe. J'ai également obtenu des cristaux de la KR<sub>1</sub> de la PKS lankamycine (une KR modèle épimérisante) afin de résoudre sa structure et d'identifier le ou les résidus catalytiques impliqués dans la réaction d'épimérisation. Cependant le problème de la phase n'a pas été résolu, ce qui n'a pas permis d'aboutir à la structure. Enfin, je me suis efforcé à identifier le domaine responsable de la stabilisation de la configuration du  $C_{\alpha}$  de la chaîne en cours de synthèse. L'objectif initial était de confirmer, que le domaine *acyl carrier protein* (ACP) n'est pas impliqué puis d'évaluer l'implication du domaine kétosynthase (au niveau d'un di-domaine kétosynthase-acyl transférase (KS-AT)) et de la KR.

J'ai également contribué à la réalisation d'un second projet, pour lequel notre équipe a mis en évidence une nouvelle classe de DD présent chez les PKS de type *trans*-AT. Les résultats montrent que l'un des DD étudié est une protéine intrinsèquement désordonnée (IDP) dont le repliement est induit par son partenaire, ce qui représente la première IDP décrite pour les protéines de métabolite secondaire bactérien. De plus, la structure RMN des deux partenaires fusionnés a mis en évidence une nouvelle topologie pour une interaction protéine-protéine. Enfin, par diffusion des rayons X aux petits angles (SAXS), nous avons pu visualiser l'interface complète entre deux sous-unités de PKS de type *trans*-AT, révélant une chambre de réaction protégée dans laquelle les chaînes de polykétide peuvent croître.

**Mots-clés :** polykétide, polykétide synthase, kétoréductase, stéréochimie, domaine de docking

Polyketides are natural products assembled by bacteria, fungi and plants which exhibit a variety of therapeutic activities, including anti-fungal, antibiotic, anti-cancer, immunosuppressant and anti-cholesterolemic properties. Given their medical and economic importance, there is significant interest in identifying new structures with novel biological activities. A promising strategy to create such analogues is to genetically engineer the enzymes responsible for synthesizing these molecules, the modular polyketide synthases (PKSs), an approach referred to as 'synthetic biology'. However, in order to increase the efficacy of this approach, we must understand in detail how the PKS multienzymes function. A number of issues remain to be clarified, including: i) how the ACP which carries the growing chain interacts with all of the catalytic domains, as well as the way in which the domains act together to control stereochemistry; and, ii) the means by which the multiple subunits present within each PKS manage to communicate with each other in an efficient and specific way; here, the focus is on the 'docking domains' which are essential to these interactions.

During my thesis, I identified two ketoreductase (KR) domains which when introduced into an intrinsically non-epimerizing modular context, were able to efficiently epimerise at the  $C_{\alpha}$  of a model polyketide. I also showed that the modular context in which the KR functions has an influence on the ultimate stereochemical outcome, which must be taken into account in future attempts to alter this critical aspect of PKS structure. I also succeeded in crystallizing a model epimerizing KR (KR<sub>1</sub> of the lankamycin PKS) with the aim of solving its structure in order to identify the catalytic residue (or residues) responsible for the epimerization. However, we could not overcome the phasing problem, which hampered resolution of the structure. Finally, I carried out experiments aiming to identify the domain within the PKS modules responsible for stabilizing the stereochemistry at the  $C_{\alpha}$  position of the polyketide chains. The initial objective was to confirm that the acyl carrier protein (ACP) domain of a model module is not implicated, and then to assess the involvement of the ketosynthase (within the context of a ketosynthase-acyl transferase (KS-AT) didomain), as well as the KR.

I also made essential contributions to a second project, in which the group identified a novel family of docking domains (DD) present in the *trans*-AT type of PKS. The available evidence argues that one of the two model DDs studied is an intrinsically disordered protein whose folding is induced by its partner, which would represent the first IDP to be described in bacterial secondary metabolism. Furthermore, the NMR structure of a fusion of the two DDs revealed a protein-protein complex of novel topology. Finally, by small-angle X-ray scattering (SAXS), we were able to visualize a complete intersubunit interface within a *trans*-AT PKS, revealing a protected reaction center in which polyketide chains can grow.

**Key words :** polyketide, polyketide synthase, ketoreductase, stereochemistry, docking domains