



HAL
open science

La biologie structurale des polykétide synthases de type trans-AT

Jonathan Dorival

► **To cite this version:**

Jonathan Dorival. La biologie structurale des polykétide synthases de type trans-AT. Biochimie [q-bio.BM]. Université de Lorraine, 2016. Français. NNT : 2016LORR0265 . tel-01527364

HAL Id: tel-01527364

<https://theses.hal.science/tel-01527364>

Submitted on 24 May 2017

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

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.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

Jonathan DORIVAL

**La biologie structurale des polykétide synthases de type
trans-AT**

Le 18 novembre 2016

Membres du jury :

Rapporteurs :	Mme Véronique RECEVEUR BRECHOT	Directrice de recherche, CNRS, Marseille
	M Peter F. LEADLAY	Professeur, Université de Cambridge, Cambridge
Examineurs :	Mme Mirjam CZJZEK	Directrice de recherche, CNRS, Roscoff
	M Benoît GUILLOT	Professeur, Université de Lorraine, Nancy
	M. Arnaud GRUEZ	Maître de conférences, Université de Lorraine, Nancy,
	Mme Kira J. WEISSMAN	Professeure, Université de Lorraine, Nancy, directrice de thèse

Je tiens tout d'abord à remercier l'ensemble du laboratoire IMoPA pour m'avoir donné l'occasion de réaliser ma thèse dans les meilleures conditions, dans un bâtiment flambant neuf, et surtout avec l'ensemble des appareils nécessaires à notre disposition.

Merci également aux membres du jury de thèse, à Peter Leadlay pour avoir accepté de rapporter mes travaux de thèse et pour vos remarques pertinentes durant la soutenance. Un immense merci à Véronique, pour tous tes conseils, remarques, d'avoir toujours cherché à me faire progresser durant chacune de nos discussions. Merci également à Mirjam d'une part pour avoir accepté de juger mon travail en tant qu'examinateur mais aussi de m'avoir offert l'opportunité de poursuivre dans cette voie en me proposant de travailler avec toi en post-doc, j'en suis honoré et je suis certain de pouvoir apprendre beaucoup de choses à tes côtés. Merci à Benoit Guillot, d'une part pour ta sympathie lors du congrès de Montréal et pour avoir accepté de juger mes travaux que ce soit dans mon comité de suivi de thèse ou dans mon comité de thèse, tes remarques ont toujours été pertinentes, de même que tes conseils.

Je remercie bien évidemment Kira Weissman d'avoir été une encadrante aussi présente, prévenante malgré tes responsabilités de professeure mondialement reconnue et de mère de trois adorables enfants. Je te suis énormément reconnaissant d'avoir toujours su trouver le temps pour chacune de nos discussions scientifiques, d'avoir toujours pris le temps de m'expliquer les choses patiemment. Ton regard sur cette thèse a largement contribué à sa réussite et je suis certain de la réussite de tous tes futurs doctorants. Je tiens à te faire savoir que ça a été un immense plaisir de pouvoir travailler avec une encadrante si compétente, mais également très accessible et agréable au quotidien.

Je remercie également Arnaud Gruez d'avoir été un très bon co-encadrant mais pas que. Merci d'avoir pris le temps de me former à la cristallisation, de m'en avoir aussi donné le goût et la patience nécessaire pour y parvenir. Merci de m'avoir transmis toutes les bases pour mener à bien des études par SAXS et de m'avoir transmis de nombreuses choses pour la cristallographie et la structurale en générale. Je te suis très reconnaissant (tiens, on dirait le Vosgien) pour ces trois ans et demi que tu as passé à me "parenter" au sein du labo, je suis certain que tes nombreux conseils, tant scientifiques que dans la vie de tous les jours, me serviront au quotidien. Et merci pour toutes ces missions synchrotron passées avec toi, les nombreux fous rires et délires que l'on a pu avoir là-bas, et un immense merci pour le congrès de Montreal avec toi, que de supers souvenirs !

Je tiens également à remercier Benjamin pour tellement de choses. Tu as déjà été une sorte de grand frère du labo (tu as d'ailleurs le même âge que mon frère). Merci pour ton humour, parfois assez cynique mais toujours tellement bon ! Merci pour toutes les discussions que j'ai pu avoir toi, d'être si ouvert, si accessible, si simple dans le bon sens du terme ! Toi et Marie-Eve avaient énormément compté pour moi et je suis certain que ce sera toujours le cas dans le futur. Merci à toi de m'avoir appris tant de choses sur la purification, sur l'informatique au sens large, sur la structurale et d'avoir pris le temps de me former à la RMN. Merci pour toutes tes bonnes idées qui m'ont souvent beaucoup aidé et merci pour tous ces moments de rigolades.

Merci également à toi Christophe, à la fois pour les discussions scientifiques, mais aussi pour les discussions personnels que l'on a pu avoir et qui m'ont énormément aidé dans mes choix. Merci également pour ton humour, mais aussi pour toutes les fois où j'ai pu t'entendre râler (notamment devant la Kyocera) ce qui m'a toujours fait sourire. Et comment ne pas te remercier pour ta contribution et ton implication énorme dans le projet LkcE. Merci également pour toutes tes idées dans les clonages qui nous ont beaucoup fait avancer !

Un grand merci à Sabrina avec qui on a formé le binôme de choc des purificateurs de l'extrême ! Merci pour les (nombreuses) purifs de module que l'on a pu faire tous les deux, pour ton aide précieuse qui m'a permis de me concentrer sur d'autres parties de ma thèse. Je te remercie également de m'avoir appris tant de choses sur l'entretien et la réparation des machines de purif, j'ai trouvé tous ces moments passionnants ! Merci aussi pour ton sourire, ta bonne humeur et ton côté "maman poule" que j'ai toujours adoré !

Je tiens à remercier également Jocelyne, dit Jojo, à la fois pour ta bonne humeur, ton énergie et pour tous les services que tu peux nous rendre au quotidien dans le labo.

Merci également à Sandrine et François de m'avoir accueilli au laboratoire et de m'avoir enseigné toutes les bases de l'enzymologie lors de mon stage de M1. Merci pour tous les conseils que vous avez pu me fournir lors de ce stage et encore au cours de ma thèse.

Merci également à Sophie et Hortense, Sophie pour tes questions et tes remarques toujours très pertinentes lors de nos séminaires qui m'ont permis de prendre pas mal de recul sur mes différentes manips et Hortense pour ta gentillesse quotidienne.

Comment ne pas remercier mes précieux frères d'armes avec qui j'ai passé tant de super moments ! A toi ma Samia, qui m'a pris sous ton bras durant mon M2 et qui m'a fait sortir de ma solitude du rez-de-chaussée. Ton humour, ton sourire, ta gentillesse et tes bons conseils ont été vraiment précieux pour moi. Je suis heureux que tu aies trouvé un super post-doc aux States !

Merci à toi mon Titi, pour tous les délires (parfois tellement débiles mais toujours aussi bons), tu es et tu resteras toujours un ami très proche ! On a partagé beaucoup de choses durant mes trois ans, y compris le sujet DD et je garde chaque moment en mémoire. Je suis heureux que tu puisses vivre ton rêve Caincain, aux pays des alligators et des cannibales ! Toi et ta femme êtes des personnes vraiment supers !

Merci à ma ptite Fanny, tu es arrivée au moment où Titi nous quittait et j'ai été hyper heureux de retrouver quelqu'un d'aussi délirant ! Merci pour tous ces supers moments !! Toutes les vanes gentilles que l'on a pu échanger, toutes ces fois où j'ai réussi à te faire râler, tous les moments où on a pu parler plus sérieusement, bref tout ça a été très précieux pour moi. Soit dit en passant (transition parfaite pour toi mon JC), une question me taraude : Elle est où la balle !?

Oh JC, my god, my brother ! Comment ne pas te remercier ! Un immense merci pour ta gentillesse, ton côté barjot que tu caches plutôt bien, pour tes innombrables conseils et les discussions enzymo que j'ai pu avoir avec toi ! Partager ces années avec toi a été un grand bonheur, mais en même temps comment ne pas être heureux d'être au côté de Dieu ! Ne change jamais ; je vous souhaite tout le meilleur à toi et Juju !

Un grand merci à toi M. Flo ! Pour tout ! A la fois ta gentillesse, tous les délires que l'on a partagé, tu es arrivé le dernier parmi la troupe des frères d'armes mais ce n'est pas pour autant que tu n'as pas eu le temps de devenir un ami tellement précieux ! Je souhaite que tu fasses une thèse super et que tu n'ais que des bonnes choses dans ta vie, tu le mérites plus que tout !

A toi M. Baloo, merci pour toutes discussions intéressantes que j'ai eu avec toi, pour être venu tous ces midi sonner l'heure du miam miam, pour ta gentillesse, tes bons conseils de bio-moi et les délires qu'on a pu partager. Bon courage pour ta thèse (eh oui, désolé je t'ai cramé en route) et je te souhaite tout le meilleur pour la suite dont je suivrais l'évolution avec plaisir !

Comment ne pas te remercier mon Beubeu ! On se suit depuis la L2 et chaque année a vraiment été merveilleuse en ta compagnie. T'es un ami comme chacun rêverait d'avoir, tu as toujours été là pour me remonter le moral à ta façon (plutôt efficace), t'es toujours partant pour aller se faire une soirée ou une sortie quelconque et ça c'est un vrai bonheur. Tu es quelqu'un sur qui j'ai toujours pu compter et je te rendrai toujours la pareille ! Bon courage pour la préparation de la soutenance. Je sais qu'on gardera toujours le contact et ma porte te sera toujours ouverte ! Ne laisse pas ma console s'ennuyer sans toi aussi ;)

Merci à toi ma ptite Daf. Ta bonne humeur, ton côté naturel et ton extrême gentillesse ont toujours compté énormément à mes yeux ! Je te souhaite tout le meilleur, et je suis assez dégouté de fuir au moment où toi, Flo et Géraud débarquez au labo ! Mais je sais qu'on ne se perdra pas de vue, loin de là. J'ai que des souvenirs merveilleux avec toi !

Un grand merci aussi à ma ptite Mahdia, à qui je souhaite tout plein de courage et tout le meilleur, à la fois dans ta thèse et dans ta vie de ptite louloute ! Ta joie de vivre a toujours été hyper réconfortante pour moi.

Un immense merci à ma super Mélu, je ne sais pas par où commencer tant j'aurais à dire mais tu sais déjà ce que tu représentes pour moi ! Ta douceur, ton sourire, ta sensibilité m'ont toujours énormément touché, et ton soutien durant cette dernière année de thèse a été très important pour moi. Je suis et je resterai toujours hyper proche de toi. Je n'ai qu'à levé mes yeux et voir le Groot posé fièrement au-dessus de mon bureau pour me souvenir de tous les supers moments passés avec toi ! Et il y en a tellement !

Je tiens aussi à remercier Alex, Mûsieur l'ingénieur en chef, avec qui chaque discussion était intéressante. J'ai adopté plusieurs de tes remarquables protocoles de production et purification de protéines et je continuerai à les propager autour de moi. Merci pour toutes tes explications techniques sur les différents appareils du labo et merci pour tous les délires, chansons, etc que l'on a pu partager ensemble.

Merci à Boubou pour ta bonne humeur, pour ta simplicité et pour tous les moments de rigolade que l'on a eu !

Merci à ma ptite Syrine, pour toutes les fois où tu as pu me faire attraper des énormes fous rires avec tes blagues à deux balles d'égyptienne ! Vous êtes géniaux G-Rhô et toi !

Je tiens à remercier également la présidente de notre association A-team of BioSE, Laurie, qui s'est investie dans le maintien du bon fonctionnement de l'asso. Merci également pour ta gentillesse, tout le soutien que tu as pu me témoigner durant ces deux dernières années de thèse.

Je veux également remercier deux personnes merveilleuses, qui ont toujours cru en moi. Merci à mes parents que j'aime de tout mon cœur. Grâce à votre bonté et votre générosité, nous n'avons jamais manqué de rien et encore moins d'amour et de soutien. Vous êtes des parents modèles et si je passe cette thèse aujourd'hui c'est en grande partie grâce à vous deux.

Je tiens également à remercier une des personnes qui compte le plus à mes yeux. A toi mon frêrot adoré, mon Alex à qui je dois tant de choses. Tu m'as toujours protégé et couvé lorsque j'étais petit et tu es toujours aussi présent maintenant. Je ne te le dis pas assez souvent mais le fait d'avoir un frère si formidable compte tellement pour moi. Tu as été un exemple pour moi et j'ai toujours tout fait pour te rendre fier de moi. A toi aussi ma ptite Cid qui partage la vie de mon frère et j'en suis particulièrement heureux, de vous voir vous entendre si bien et d'avoir une belle sœur aussi adorable !

Je tiens bien évidemment à remercier celle qui m'a le plus encouragé, et avec qui je partage désormais mon quotidien avec beaucoup de joie. Merci à toi Anaïs, ma chérie de tout ce que tu peux m'apporter comme soutien, amour et réconfort et merci à toi de croire aussi fort en moi. J'essaierai d'être à la même hauteur pour t'encourager également quand ton tour sera venu de finir ta thèse.

Un immense merci à mes beaux-parents, pour tous ces bons repas chez vous, pour votre gentillesse, et pour tout ce que vous avez pu faire pour nous deux. Un gros merci à ma ptite sœur, ma ptite Ambre, ta présence à ma thèse m'a beaucoup touché et tu es quelqu'un de super pour qui je serais présent le plus possible !

Je remercie bien évidemment le reste de la petite famille, à mes cousins et cousines qui mes sont tellement chers, Samsou et Laurie, Didine et Bilou, ma Jod, Pascaline, Romu, Nini et Amadou, ma ptite Marie, Simon, Maëlle et Arthur. Un grand merci également à mes oncles et tantes qui m'ont chouchouté durant toutes ces années.

Un immense merci à ma Mémé qui a toujours été présente pour moi et qui nous a toujours tant donné malgré le peu de moyens que vous aviez, merci pour la maison de Gruissan dans laquelle on a tous pu passer des supers vacances et merci pour tous ces merveilleux moments en famille et tous ces souvenirs. J'ai une énorme pensée pour mon pépé, qui je suis sûr doit être fier de son petit-fils de là où il est. Sache en tout cas que je pense très souvent à toi et que ton modèle me donne la force d'aller de l'avant !

Je profite de cette occasion pour remercier ma Soprane, ma merveilleuse amie. On veille l'un sur l'autre depuis que l'on se connaît. Il y a eu des bons moments, d'autres moins mais ils sont clairement effacés par tout ce que j'ai pu vivre de merveilleux en ta présence. Ton adorabilité, ta joie de vivre, ton innocence, ton caractère (tantôt de cochon, tantôt d'ange), ton côté taquine, tout ça contribue à faire de toi quelqu'un en qui j'ai une entière confiance et de qui je suis extrêmement proche ! Merci pour tous ces délires, ces moments de complicité avec toi et merci de ton soutien et de remonter le moral dans mes moments de doute !

Merci également à Chiraz, mon petit bout, une amie qui m'a encouragé plus fort que quiconque, l'une des personnes à me comprendre le mieux. On a veillé l'un sur l'autre depuis le jour où je t'ai rencontré. Tu es mon ange gardien et tes conseils sont d'une importance capitale ! Merci pour tous ces bons moments, merci de m'avoir fait découvrir l'Algérie et de m'avoir ouvert les yeux sur tant de choses. Ta rencontre a changé ma vie et j'espère pouvoir être là quand tu auras besoin pour rendre la tienne toujours meilleure.

Merci à Djihed, ma doudou adorée, pour toute ta douceur, ton humour, ton caractère en or, bref, tu sais à quel point tu comptes pour moi !

Merci à ma ptite Camille, mon poussin, pour son extrême gentillesse, sa patience, son soutien indéfectible, pour tous les mots gentils que tu as pu me dire. Ta sincérité, ton innocence et ta folie (si proche de la mienne) te rendent tellement unique.

Merci à Beren qui, même si on ne se voit pas beaucoup, est une amie vraiment précieuse. Ta générosité, ton caractère en or, tes sourires si sincères, toutes tes qualités font de toi quelqu'un que j'apprécie et à qui je tiens très fort ! Ne doutes jamais de toi, vois le bon côté des choses, un verre est souvent plus plein qu'il n'est vide.

Un grand merci également à ma Lili, pour ta gentillesse et toutes ses fois où tu es là pour moi !

Et un immense merci pour ma ptite Kmi, ma ptite sœur de cœur, toujours là avec les mots juste pour me faire rire, me remotiver ou me donner plein de tendresse ! Tu es juste géniale, ne change jamais !!!

Cette liste est bien sûr incomplète et je tiens à remercier toutes les personnes qui ont contribué à rendre cette expérience aussi enrichissante sur le plan humain que sur le plan scientifique.

Quelques proverbes bretons appris durant les premiers jours de post-doc :

Horizon pas net, reste à la buvette. Qui trop écoute la météo, passe sa vie au bistrot.
Celui qui marche droit trouve toujours la route assez large.

ABBREVIATIONS

Abréviations

6-dEB : 6-déoxyérythronolide B

A : adénylation

ACP : acyl carrier protein

ADN : acide désoxyribonucléique

AT : acyltransférase

ATP : adénosine triphosphate

β -ME : β -mercaptoéthanol

C : condensation

Cy : domaine d'hétérocyclisation

CD : dichroïsme circulaire

CoA : coenzyme-A

Cryo-EM : cryo-microscopie électronique

Da : dalton

DATD : N,N' diallyltatradiamide

DEBS : 6-déoxyérythronolide B synthase

DD : domaine de docking

DH : déshydratase

DLS : diffusion dynamique de la lumière

DMSO : diméthylsulfoxyde

DO : densité optique

E : épimérase

EDTA : acide éthylène diamine tétra-acétique

EMAA : éthyl-méthyl-acétoacétate

ER : énoyl réductase

FAD : flavine adénine dinucléotide

FAS : acide gras synthase

FMN : flavine mononucléotide

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

HC : hétérocyclase

HCS : hydroxyméthylglutaryl-Coenzyme A synthase

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

HMG : hydroxyméthylglutaryle

HR : hautement réductrice

IDP : protéine intrinsèquement désordonnée

IMAC : chromatographie type chélation de métaux immobilisés

ITC : titration calorimétrique isotherme

IPTG : isopropyl β -D-1-thiogalactopyranoside

KR : kétéoréductase

KS : kétéosynthase

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

MAD : dispersion anormale à multiple longueur d'onde

MAS : synthase de l'acide micocérosique

MAT : malonyl-acétyl transférase

MCAT : malonyl-CoA transférase

MDR : déshydrogénase/réductase à courte chaîne

mFAS : synthèse d'acide gras de mammifère

MT : méthyltransférase

NADPH : nicotinamide adénine dinucléotide phosphate

NCS : symétrie non cristallographique

NR : non réductrice

NRPS : synthèse de peptides non ribosomiques

PCP : peptidyl carrier protein

PCR : réaction de polymérisation en chaîne

PDB : banque de donnée des protéines

PEG : polyéthylène glycol

PdI : polydispersité

PKS : polykétide synthase

PPTase : phosphopantéthéinyl transférase

PUFA : acide gras polyinsaturé

PR : partiellement réductrice

RMN : résonance magnétique nucléaire

SAD : diffusion 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

SLS : diffusion statique de la lumière

Sfp : surfactine phosphopantéthéinyl-transférase

TCEP : tris(2-carboxyethyl)phosphine

TE : thioestérase

TFA : acide trifluoroacétique

TFE : trifluoroéthanol

T_m : température d'hybridation

SOMMAIRE

Introduction	1
I. Les métabolites secondaires	1
1. Généralités et classification	1
2. Les polykétides	3
A) Des métabolites secondaires à forte valeur ajoutée	3
B) De la découverte des polykétides à l'étude des polykétide synthases	6
II. Les fatty acid synthases, un ancêtre commun partagé avec les PKS	9
1. Les FAS de type I d'origine fongique	9
2. Les FAS animales, des synthases homologues aux PKS	13
III. Les PKS	17
1. Organisation en cluster	17
2. Les PKS de type III	18
3. Les PKS de type II :	19
4. Les PKS de type I	20
A) Les PKS de type I itératives	22
B) Les PKS de type I modulaires	22
a. Les PKS de type I modulaires de type <i>cis</i> -AT	22
b. Les PKS de type I modulaires de type <i>trans</i> -AT	24
IV. L'attractivité des PKS de type I pour la biologie synthétique	29
1. Les PKS de type <i>cis</i> -AT, un modèle pour l'approche d'ingénierie	30
2. Les PKS de type <i>trans</i> -AT : une alternative intéressante pour créer de la diversité	35
3. Une approche d'avenir, mais de nombreux points restent à éclaircir	36
V. Etat des connaissances des structures et mécanismes des domaines PKS	37
1. Les domaines « acyl carrier protein »	37
2. Les domaines acyltransférase	38
3. Les domaines kétosynthase	41

4. Les domaines kétoréductase _____	44
5. Les domaines déshydratase _____	46
6. Les domaines énoyl réductase _____	48
7. Les domaines méthyltransférase _____	50
8. Les domaines thioestérase _____	51
VI. Organisation structurale des modules PKS _____	53
1. La structure de la mFAS comme modèle pour les PKS de type I ? _____	54
2. Structure cristallographique d'une PKS itérative _____	57
3. Etat des connaissances sur les réarrangements conformationnels associés à la catalyse _____	58
A) Structure d'un module entier par cryo-microscopie électronique (cryo-EM) _____	59
B) Organisation structurale d'un module et bi-module par SAXS _____	63
VII. Les domaines de docking : gardiens de la continuité et de l'ordonnement de la synthèse des polykétides _____	64
1. Les domaines de docking de type 1a _____	66
2. Les domaines de docking de type 1b _____	68
3. Les domaines de docking de type 2 _____	70
4. Une classe à part de domaines de docking _____	71
VIII. Les enzymes de maturation post-PKS : une cible de choix pour l'ingénierie ? _____	73
IX. Choix de deux systèmes modèles de PKS de type trans-AT _____	75
1. La virginiamycine M et sa PKS _____	75
2. La PKS synthétisant la lankacidine C. _____	79
Objectifs _____	84
Matériel et méthodes _____	86
I. Matériel biologique _____	86
1. Souches bactériennes utilisées _____	86

2. Vecteurs utilisés _____	87
3. Milieux de culture _____	87
4. Solutions d'usage courant _____	90
II. Sous-clonage, production et purification des protéines recombinantes	90
1. Sous-clonage de la séquence codant la LkcE _____	90
2. Mutagenèse dirigée _____	91
3. Production et purification de la LkcE et de ses mutants _____	92
A) Culture permettant l'expression de la LkcE _____	92
B) Lyse cellulaire et purification de la LkcE _____	93
a. Lyse cellulaire _____	93
b. Chromatographie d'affinité Nickel-Sépharose _____	93
c. Protéolyse de la protéine de fusion et du tag histidine suivie d'une chromatographie d'affinité Nickel-Sépharose inverse _____	94
d. Chromatographie échangeuse d'anions _____	94
e. Chromatographie d'exclusion stérique _____	94
4. Production de la SeLkcE _____	95
III. Production, extraction et purification du substrat de la LkcE.	95
IV. Etudes préliminaires de la LkcE	98
1. Homogénéité en solution _____	98
2. Vérification de l'intégrité de la LkcE et caractérisation du co-facteur flavinique par spectrométrie de masse _____	98
V. Etude de la LkcE par cristallographie aux rayons X	98
1. La cristallisation de la LkcE _____	99
A) Principes et techniques de cristallisation _____	99
B) Recherche et amélioration des conditions de cristallisation de la LkcE _____	103
a. Criblage de conditions de cristallisation _____	103
b. Amélioration des cristaux _____	104
c. Trempage par diffusion des substrats _____	105
2. Congélation et enregistrement des données de diffraction _____	105
A) Congélation des cristaux _____	105

B)	Enregistrement des données de diffraction	106
3.	Traitement des données de diffraction et phasage	107
A)	traitement des données	107
B)	Résolution du problème de phase	108
a.	Phasage par diffusion anormale	108
i.	La diffusion anormale simple (SAD)	108
ii.	Le MAD	110
iii.	Phasage expérimental de la LkcE	111
b.	Phasage par remplacement moléculaire	111
4.	Construction et affinement des structures	112
VI.	Test d'activité et détermination des paramètres cinétiques de la LkcE	114
1.	Test d'activité de la LkcE avec analogues de substrat	114
2.	Test d'activité de la LkcE avec son substrat naturel (LC-KA05)	115
VII.	La diffusion des rayons X aux petits angles (SAXS)	115
1.	Principe et acquisition des données	115
1.	Traitement des données	118
A)	Loi et Représentation de Guinier	118
B)	Fonction de distribution de distance	120
C)	Représentation de Kratky	121
2.	Interprétation des données SAXS et modélisation structurale	122
A)	Calcul d'enveloppe <i>ab-initio</i>	123
B)	Comparaison de structures tout atome aux données SAXS	124
C)	Modélisation en corps rigide	125
D)	Détermination de la masse moléculaire sans utilisation du I(0)	126
	Résultats	127
I.	Structure en solution d'un module de PKS trans-AT : un premier pas dans la compréhension de la communication intra-modulaire	127
II.	Caractérisation d'une nouvelle classe d'interface de docking chez une PKS de type trans-AT	175

III. Etude d'une enzyme post-PKS impliquée dans la cyclisation de la lankacidine C	212
1. Analyse <i>in-silico</i> de la LkcE	212
2. Clonage et purification de la LkcE et des mutants	214
A) Clonage de la séquence codant la LkcE	214
B) Purification de la LkcE	214
C) Homogénéité en solution	215
D) Vérification de l'intégrité de la LkcE _{wt} et caractérisation de la nature du co-facteur flavinique	216
3. Etude structurale de la LkcE	218
A) Criblage des conditions de cristallisation de la LkcE _{wt} et cristallisation de la Se-LkcE	218
B) Phasage SAD et construction du premier modèle	221
C) Cristallisation de la LkcE avec analogues de substrat et construction du modèle	223
a. Obtention de cristaux monocliniques	223
b. Obtention de cristaux tétragonaux pour la protéine native	226
D) Caractérisation structurale du mutant R326Q de la LkcE trempé avec le LC-KA05	230
a. Cristallisation des mutants, collecte de données et affinement de la structure du mutant R326Q en complexe avec le substrat natif	230
b. Analyse de la structure	234
E) Co-cristallisation de chaque mutant et l'enzyme wt avec le LC-KA05 afin d'obtenir le produit dans le site actif	240
4. Mise au point du test d'activité de la LkcE et de la détermination des paramètres cinétiques	240
A) Test d'activité de la LkcE avec les analogues de substrat (DATD et EMAA)	240
B) Test d'activité de la LkcE avec le substrat naturel	245
Discussion et perspectives	251
Références bibliographiques	267

INTRODUCTION

I. Les métabolites secondaires

1. Généralités et classification

Au contraire des métabolites primaires qui sont cruciaux pour les fonctions vitales de la cellule, les métabolites secondaires sont des composés non essentiels aux bactéries, champignons, plantes et animaux qui les produisent. Toutefois, bien que leur absence ne soit pas létale, ces composés peuvent conférer des avantages non négligeables et favoriser la survie ainsi que la croissance des espèces les produisant, augmentant ainsi la compétitivité de l'organisme hôte au sein de son biotope (Roze, *et al.*, 2011). En effet, la production de ces métabolites secondaires apporte différents avantages à l'organisme les synthétisant. Dans la plupart des cas, ces métabolites secondaires pourraient servir à éliminer les autres espèces partageant le même biotope que l'espèce productrice (Demain, et Fang, 2000). Ils peuvent également agir en tant que transporteur de métaux (sidérophores) et servir d'antibiotiques en privant les autres espèces de fer, ou pour maintenir la balance Na^+/K^+ (ionophores) (Kanne, et Zahner, 1976). Enfin, ils pourraient être impliqués dans une communication inter-espèces, permettant l'établissement de symbiose entre plusieurs organismes (Paul, *et al.*, 1981).

Bien qu'ils soient principalement issus de quelques métabolites primaires bien particuliers (l'acétyl-CoA et acides aminés majoritairement), il existe une grande variété de métabolites secondaires. Ainsi, à ce jour, plus d'un million de métabolites secondaires ont été découverts au sein de la biosphère et une grande quantité reste encore à découvrir (Demain, 2014).

Ces composés ont été classés en six grandes catégories (Agostini-costa, *et al.*, 2012) :

1) Les alcaloïdes, qui sont des métabolites pour la plupart d'origine végétale, dérivées d'acides aminés, majoritairement cycliques et composés d'azote. Certaines molécules de cette famille ont des applications médicales comme la morphine, et d'autres peuvent être couramment consommés dans la vie de tous les jours : nicotine et caféine, par exemple.

2) Les terpènes, qui sont majoritairement produits par les plantes et les champignons. Ce sont des composés synthétisés à partir d'unités de base à 5

carbones, l'isoprène, dérivant de l'isopentényl pyrophosphate (IPP) formé à partir d'acétyl-CoA. Parmi les terpènes, on retrouve des composés comme le β -carotène, le limonène, mais aussi des molécules à fort intérêt pharmacologique comme le paclitaxel utilisé comme anticancéreux (Taxol®), ou bien encore industriel comme le caoutchouc. Les stérols, produits par de nombreux organismes, sont également classés dans cette famille.

3) Les composés phénoliques principalement dérivés du catabolisme de la phénylalanine. Cette famille regroupe une grande quantité de métabolites secondaires assez différents, parmi lesquels on retrouve les tannins, les coumarines et certains flavonoïdes.

4) Les hétérosides, également appelés glycosides, qui sont des composés issus de la condensation d'un ose avec une fraction aglycone de nature diverse (issue de terpènes, de composés phénoliques, de composés soufrés...).

5) Les acides aminés et peptides non protéinogènes produits par certaines légumineuses et qui peuvent se révéler toxiques à forte concentration. En effet ces composés ont une structure proche des acides aminés protéinogènes et peuvent interférer avec ceux-ci et créer des désordres physiologiques. Dans cette catégorie sont également classés les peptides issus de synthétases de peptide non ribosomal (NRPS) qui peuvent incorporer ou non des acides aminés non protéinogènes.

6) Et enfin, les polykétides (Agostini-costa, *et al.*, 2012) (Figure 1).

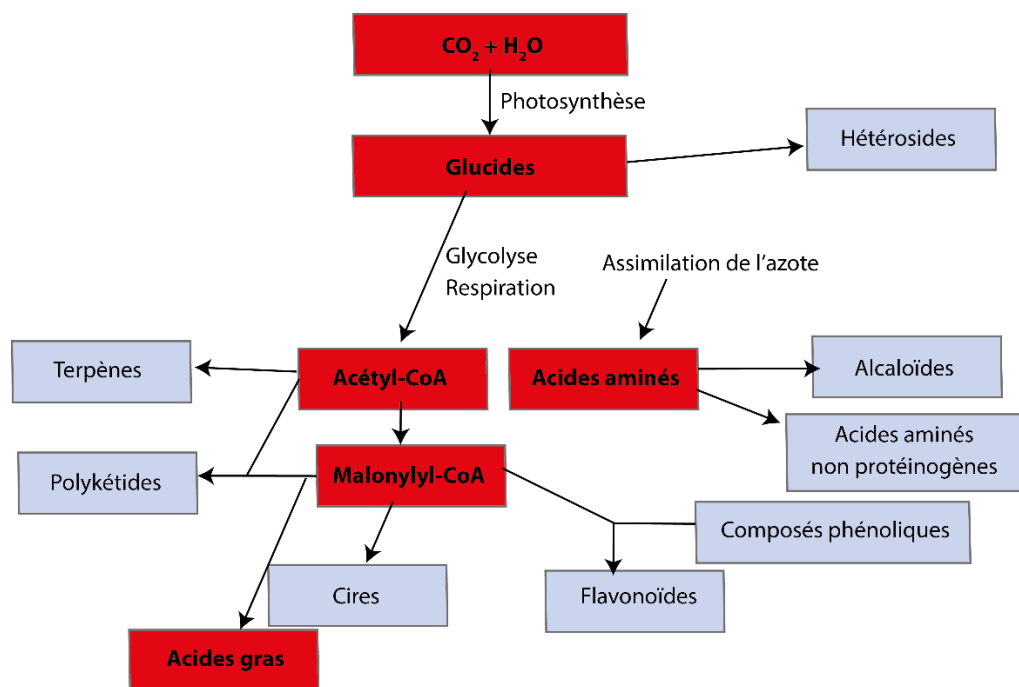


Figure 1 : Relation entre métabolisme primaire et métabolisme secondaire.

Les métabolites primaires sont mis en évidence par un encadré rouge alors que les métabolites secondaires sont indiqués sur fond gris. Ceci représente bien le lien étroit qu'il existe entre le métabolisme primaire, essentiel à la vie de l'organisme, et le métabolisme secondaire qui offre des avantages à la survie, bien que la frontière entre les deux métabolismes soit parfois assez mince.

2. Les polykétides

A) Des métabolites secondaires à forte valeur ajoutée

Les polykétides constituent une large classe de métabolites secondaires présentant une grande diversité, que ce soit en termes de structure ou de fonction. Leur fonction biologique chez les organismes producteurs reste encore assez vague, bien qu'il soit supposé qu'ils aient des rôles défensifs. Certains polykétides ont des rôles plus manifestes, comme ceux produits par des amibes sociales qui servent dans la médiation des signaux, alors que ceux produits par *Mycobacterium tuberculosis* favorisent la formation de la paroi cellulaire (Hertweck, 2009). De même, certains polykétides constituent des toxines et servent de facteurs de virulence pour les espèces qui les produisent (Stinear, *et al.*, 2004 ; Hong, *et al.*, 2005)

Les polykétides constituent des composés à très forte valeur ajoutée. En effet, nombre de ces composés sont utilisés en tant qu'antibiotiques (rifamycine,

érythromycine A et virginiamycine M, par exemple), anticancéreux (lankacidine, épothilone B, doxorubicine), antifongiques (amphotéricine), immunosuppresseurs (rapamycine) ou bien encore antiparasitaires (ivermectine). Leur synthèse est donc clairement un enjeu capital pour les industries pharmaceutiques mais aussi agrochimiques. Ainsi, les polykétides représentent chaque année plus de 20 milliards de dollars de chiffre d'affaire (Weissman, et Leadlay, 2005).

En plus de leur grande diversité fonctionnelle, les polykétides présentent une grande variété de structures (Figure 2). Ces molécules partagent comme traits communs de constituer des poly-cétones, et leur biosynthèse, que nous détaillerons par la suite. Certains de ces composés forment une structure très rigide composée de nombreux cycles aromatiques alors que d'autres polykétides sont substantiellement plus réduits, étant juste bloqués dans un nombre restreint de conformations par une cyclisation entre les deux extrémités (Weissman, 2009). Cette complexité, ainsi que la présence de très nombreux groupements fonctionnels et d'une stéréochimie bien définie, constituent à la fois un avantage, puisque cela confère à ces molécules de fortes et très nombreuses activités pharmacologiques, mais aussi un inconvénient de taille, puisque les polykétides sont des molécules complexes et leur synthèse chimique se révèle laborieuse.

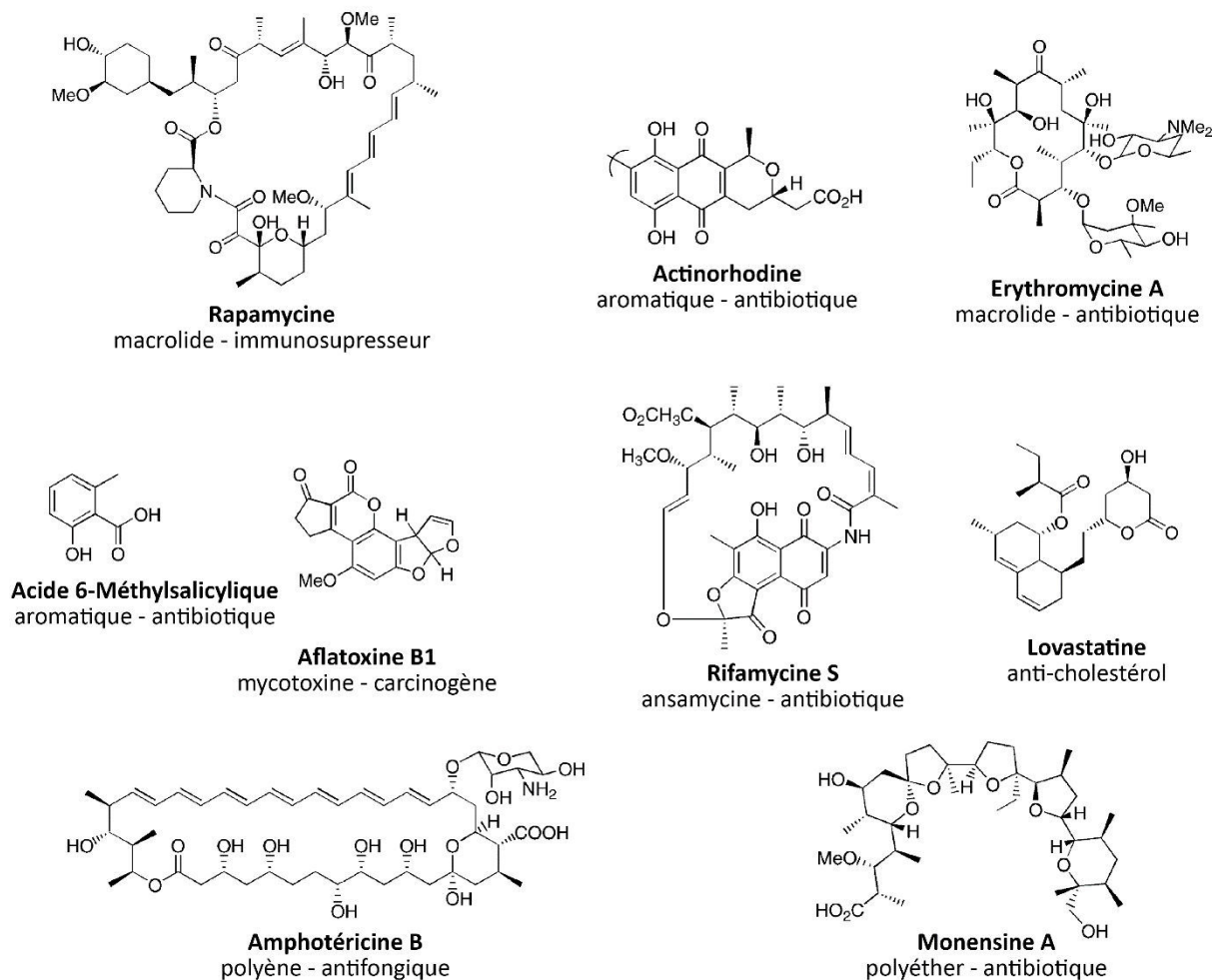


Figure 2 : Exemple de polykétides (adapté de Staunton et Weissman, 2001)

Cette figure montre la grande diversité de structures et de fonctions des polykétides. Le nom du polykétide est indiqué en gras et le type de polykétide ainsi que sa fonction biologique sont indiqués en dessous.

Ainsi la synthèse de ces composés représente un challenge pour le domaine médical. La découverte de nouvelles structures de polykétides au sein de la biosphère constitue la stratégie la plus simple pour disposer de nouvelles molécules à fort potentiel pharmacologique (Bachmann, *et al.*, 2014). Cependant il existe d'autres stratégies comme l'hémi-synthèse chimique. Ceci consiste à modifier chimiquement des composés déjà existant, ce qui permet de modifier de façon rationnelle certaines molécules afin d'optimiser leur activité. Deux dérivés de l'érythromycine A, le Zithromax et la Biaxine, ont été développés *via* cette méthode (Peltier, 1992). Une autre stratégie prometteuse consiste en l'ingénierie génétique des voies de biosynthèse des polykétides. Cette approche, désignée sous le terme de biologie

synthétique, a pour but de modifier les enzymes produisant les polykétides : les polykétide synthases (PKS) (Weissman, 2016). Cette stratégie est apparue dans le début des années 90, avec la découverte des clusters de gènes codant les PKS (Cortés, *et al.*, 1990) et le développement des techniques de séquençage et l'annotation de génomes. Cependant, l'efficacité de cette approche dépend d'une connaissance accrue des systèmes enzymatiques produisant ces molécules d'intérêt.

B) De la découverte des polykétides à l'étude des polykétide synthases

En 1893, à l'Université de Londres, James Collie découvre une série de réactions qui mène à la synthèse d'un composé aromatique appelé orcinol. Il décrit alors la formation de ce composé, à partir d'acide déhydroacétique, et ce mécanisme passe par un intermédiaire qualifié de poly-cétone (Collie, et Myers, 1893) (Figure 3).

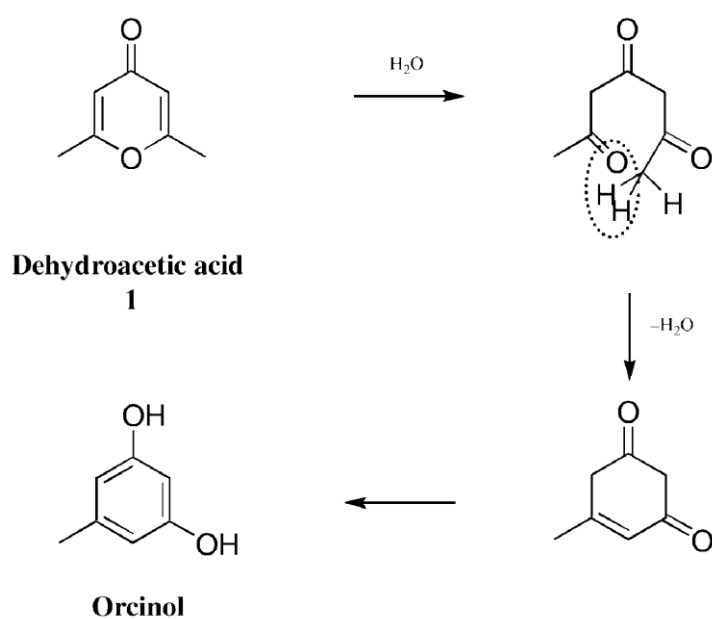


Figure 3 : Schéma de la synthèse d'orcinol proposé par Collie. (Staunton, et Weissman, 2001)

L'ajout d'une molécule d'eau sur l'acide déhydroacétique (1) entraînerait l'ouverture du cycle et la formation d'un intermédiaire poly-cétone (2). Le relargage de la molécule d'eau provoquerait la reformation du cycle, pour ensuite former l'orcinol (3)

En 1917, Robinson propose un mécanisme similaire pour la synthèse de tropinone (Robinson, 1917). Dans son livre de 1955, il développe aussi l'idée que les polyphénols peuvent être produits à partir de poly-cétones.

Dans les années 50, Birch permet une avancée considérable de ces études puisqu'il avance l'hypothèse que les polykétides peuvent être produits à partir d'unités acétate par des condensations successives, et il vérifie sa théorie en administrant des

précurseurs marqués au ^{14}C à un organisme qu'il pense producteur de polykétides : *Penicillium patulum*. A partir d'acétate marqué au ^{14}C sur la position 1, il parvient à obtenir l'acide 6-méthyl-salicylique, le polykétide attendu, avec 4 atomes radioactifs (qu'il parvient à identifier grâce à une décomposition chimique du composé). Ce qui valide son hypothèse, puisque quatre unités acétate sont nécessaires pour former ce composé.

A partir des années 60, l'émergence de la RMN et de la spectrométrie de masse ont permis la compréhension de la biosynthèse de nombreux autres polykétides. En effet, jusqu'à cette époque, les études de formation de produit ne se faisaient que par dégradation chimique pour produire des fragments identifiables, qui étaient ensuite assemblés de façon fictive pour imaginer un mécanisme de synthèse du composé étudié. Dès lors, le marquage spécifique de précurseurs au ^{13}C et ^2H a permis d'élucider le mécanisme de biosynthèse de nombreuses molécules, comme l'érythromycine A, par exemple.

Plus tard, avec l'essor de la génétique, il a été possible de rattacher la synthèse de ces polykétides à la présence de clusters de gènes codant plusieurs enzymes identifiées ultérieurement comme PKS. En effet en 1979, Rudd et Hopwood ont réalisé une série de mutations dans le génome de *Streptomyces coelicolor* et ont identifié celles qui affectent la synthèse de l'actinorhodine, un polykétide qui sert d'antibiotique mais aussi d'indicateur de pH du fait de sa coloration pH-dépendante (Rudd, et Hopwood, 1979). Ainsi, les mutations qui sont à l'origine de la disparition de la couleur bleue des colonies dû à la synthèse d'actinorhodine ont permis d'identifier la partie génique responsable de la synthèse du polykétide. Une série de 76 mutations contiguës a été effectuée dans cette partie génique et a ensuite été suivie d'un classement des mutants en six classes phénotypiques différentes basées sur l'activité antibiotique, l'accumulation de précurseurs colorés due à la formation de produits abortifs et la complémentarité entre plusieurs mutants pour rétablir la synthèse d'actinorhodine. Ces mutations localisés dans le même cluster a permis de mettre en évidence la présence d'au minimum six gènes intervenant dans la synthèse du polykétide.

Plus tard, le génome de *S. coelicolor* sera séquencé et ce cluster de gènes sera annoté par analogie avec les Fatty Acid Synthases (FAS) (Figure 4). Celle-ci partage

beaucoup d'homologies avec les PKS, que ce soit au niveau de la chimie des réactions (une succession de condensation d'unités simples aboutissant à la synthèse d'une molécule de plus haut poids moléculaire), ou de par les activités enzymatiques caractéristiques de chacune de ces deux synthèses, que ce soit les enzymes catalysant la condensation ou celles catalysant la modification des acides gras ou des polykétides.

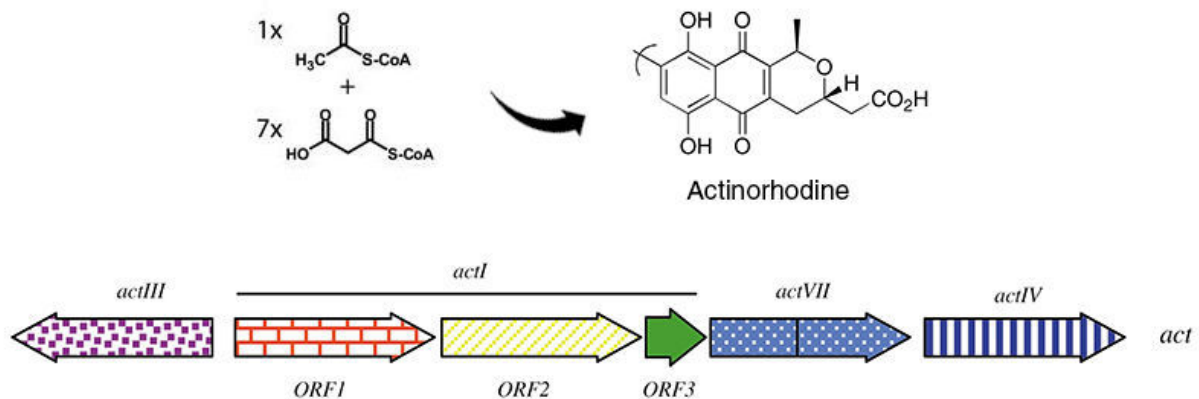


Figure 4 : Cluster de gènes codant la PKS responsable de la synthèse de l'actinorhodine. adapté de Staunton, et Weissman, 2001.

Ici sont représentés les 6 ORF identifiés par Rudd et Hopwood codant la PKS. Les ORF 1, 2 et 3 (*actI*) codent pour les enzymes responsables des réactions de condensation, à partir d'un acétyl-CoA et de 7 malonyl-CoA. Les gènes *actIII*, *actVII* et *actIV* codent pour des enzymes de modification et de cyclisation du polykétide.

Vers la fin des années 80, deux laboratoires différents cherchent à identifier les gènes responsables de la synthèse d'érythromycine A par *Saccharopolyspora erythraea*. Pour cela le groupe de Peter Leadlay va séquencer les régions proches du gène *ermE*, responsable de la résistance à l'érythromycine, puisqu'ils sont persuadés que les gènes de biosynthèse de l'antibiotique sont situés dans le même cluster que les gènes de résistance à ce même antibiotique. Ainsi ils séquencent cette région de l'ADN et recherchent des similarités avec les FAS ou d'autres PKS (J. Cortés, *et al.* 1990). Une autre équipe, celle de Léonard Katz des Laboratoires Abbott, ont reproduit ce qui a été fait sur l'actinorhodine, et ont réalisé une série de mutations dans le génome dont certaines affectent la synthèse du 6-déoxyérythronolide B (6-dEB). Ils ont donc également séquencé les régions proches de ces mutations situées non loin du gène *ermE* (Tuan, *et al.*, 1990). Au final, les deux équipes parvinrent à séquencer le cluster de gène codant la PKS synthétisant l'érythromycine A.

Comme nous l'avons vu précédemment, les PKS et les FAS sont deux synthèses qui partagent énormément de similitudes et de ce fait, bon nombre d'études sur les PKS se sont appuyées sur des résultats obtenus précédemment sur les FAS. Ces similarités en termes de mécanisme de biosynthèse, corrélées à une forte homologie de séquence primaire, démontrent que les FAS animales et les PKS de type I ont une origine évolutive commune (Jenke-Kodama, *et al.*, 2005).

II. Les fatty acid synthases, un ancêtre commun partagé avec les PKS

Les FAS ont été découvertes dans les années 60. On distingue deux types de FAS sur la base de leur origine et de leur fonctionnement. En effet, les FAS d'origine animale et fongique (type I) sont des protéines de haut poids moléculaire, composées de plusieurs domaines liés covalamment. Chaque domaine est responsable d'une activité catalytique propre lors de la synthèse des acides gras. Ils sont à distinguer des FAS d'origine bactérienne et chloroplastique (type II) qui sont composées de protéines discrètes agissant tour à tour dans la synthèse de l'acide gras.

Les FAS catalysent la synthèse d'acide gras par polymérisation successive en ajoutant de façon itérative un groupement malonyle, provenant du malonyl-CoA, sur une unité de base qui est l'acétyle. Elles ont donc un rôle prépondérant au sein des cellules, puisque les acides gras ainsi synthétisés vont servir en tant que constituant membranaire, réserve d'énergie, ou encore comme second messenger (Maier, *et al.*, 2010).

1. Les FAS de type I d'origine fongique

Chez les FAS de champignons, tous les domaines catalytiques nécessaires à la synthèse d'acides gras saturés sont présents sur deux chaînes polypeptidiques. Dans un premier temps, le groupement acétyle, l'unité de départ provenant de l'acétyl-CoA, est transféré sur le domaine « acyl carrier protein » (ACP). Ce transfert est catalysé par le domaine acyltransférase (AT). Ensuite l'ACP transfère l'acétyle sur le domaine kétosynthase (KS). Dès lors, le domaine malonyl/palmitoyl transférase (MPT) charge une unité d'extension provenant du malonyl-CoA sur l'ACP. Le domaine KS catalyse

alors la décarboxylation du malonyl et la condensation des deux unités pour former un acétoacétyl-ACP. Ensuite l'ACP apporte l'acétoacétyl à trois domaines responsables de la modification de l'état d'oxydation du carbone β . Tout d'abord, le domaine kétoréductase (KR) catalyse la réduction du groupement cétone en alcool, par oxydation concomitante du NADPH. S'ensuit une déshydratation catalysée par le domaine déshydratase (DH) ce qui entraîne la formation d'une double liaison entre les carbones α et β . Enfin, le domaine énoyl-réductase catalyse la réduction de l' α,β -énoyl en acyl saturé (Figure 5).

Comme dit précédemment, les FAS de type I fongiques ont la particularité d'être constitués de deux chaînes polypeptidiques : la chaîne β de 230 kDa qui arbore les domaines AT, ER, DH et la majeure partie du domaine splitté MPT, et la chaîne α de 210 kDa qui comporte en N-terminal l'autre partie du MPT, le domaine KR, le KS et le domaine phosphopantéthéinyle transférase (PPT). Ce dernier domaine de modification post-traductionnelle ajoute un groupement phosphopantéthéine sur la sérine canonique de l'ACP à partir du CoA. C'est ce groupement prosthétique qui va porter la chaîne en élongation et les unités d'extension tout au long de la catalyse (Fichtlscherer, *et al.*, 2000).

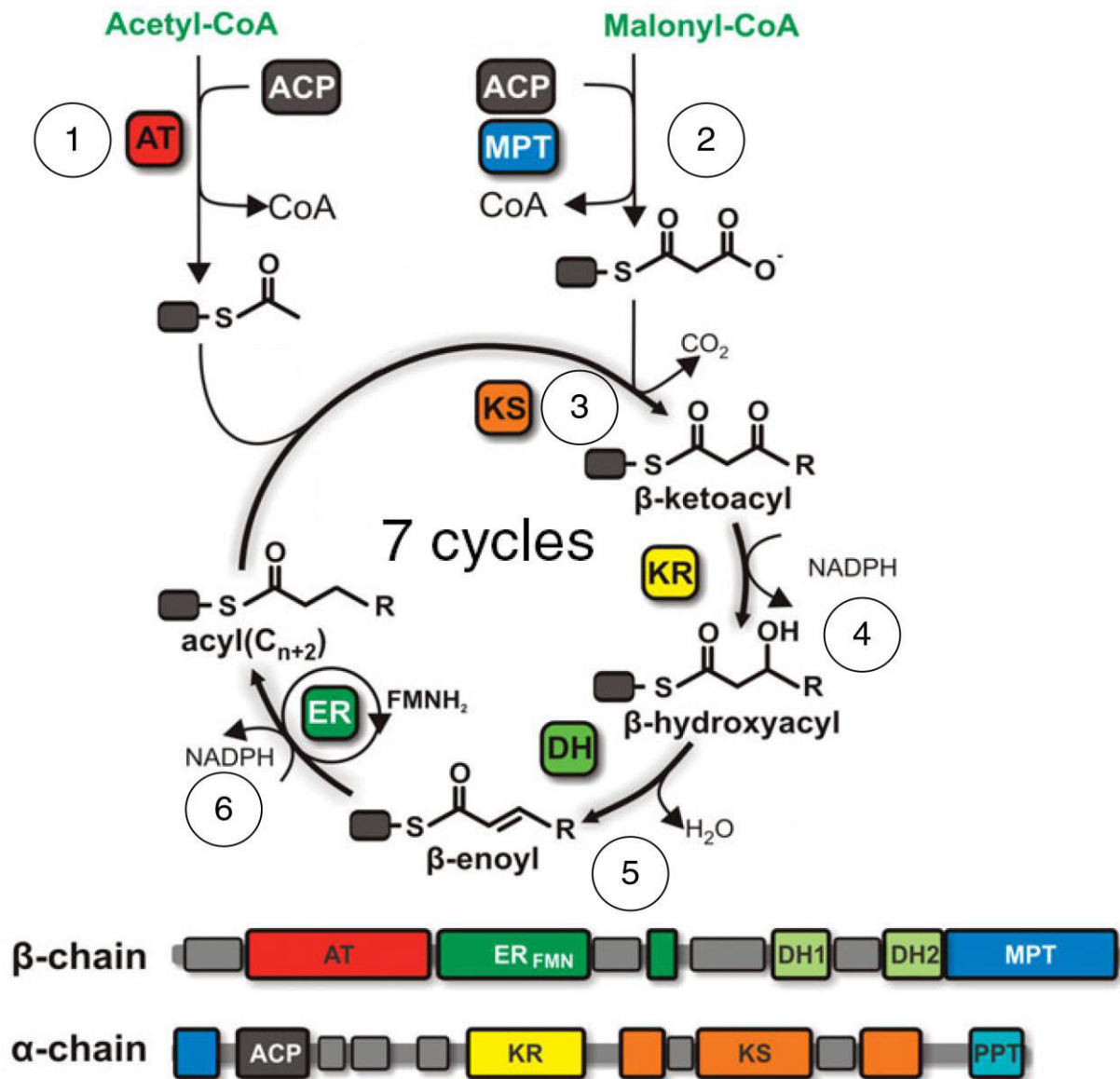


Figure 5 : Schéma réactionnel de la FAS fongique (adapté de Maier, et al., 2010).

Tout d'abord le domaine AT sélectionne l'unité de départ et la transfère sur le domaine ACP (1). Cette unité de base est ensuite transférée au niveau du site actif du domaine KS puis le domaine MPT charge l'unité d'extension sur l'ACP (2). Le KS catalyse la réaction de condensation (3), puis les domaines réducteurs agissent au niveau du carbone β pour catalyser la transformation de la cétone en acyl saturé (4, 5 et 6). Les étapes 2, 3, 4, 5 et 6 sont ensuite répétées 6 fois supplémentaires. Enfin, le C₁₆ ainsi formé est transféré sur le CoA par le domaine MPT afin de former du palmitoyl-CoA. En dessous est représentée la composition en domaines des deux chaînes polypeptidiques constituant la FAS de champignons.

La structure de la FAS de champignon a été résolue en 2007, après de longues années à essayer d'améliorer la diffraction des cristaux, qui diffractaient initialement autour de 8 Å de résolution, et à essayer de résoudre le problème de phase. Ce sont finalement les structures de la FAS de *Saccharomyces cerevisiae* et de *Thermomyces lanuginosus* qui ont été résolues à environ 3 Å de résolution (Jenni, *et al.*, 2007 ; Leibundgut, *et al.*, 2007 ; Lomakin, *et al.*, 2007).

La structure de la FAS est un hétéro-dodécamère qui peut être divisé en deux parties distincts : un anneau central formé par un hexamère de chaîne α et deux trimères de chaîne β de part et d'autre de l'anneau central (Figure 6).

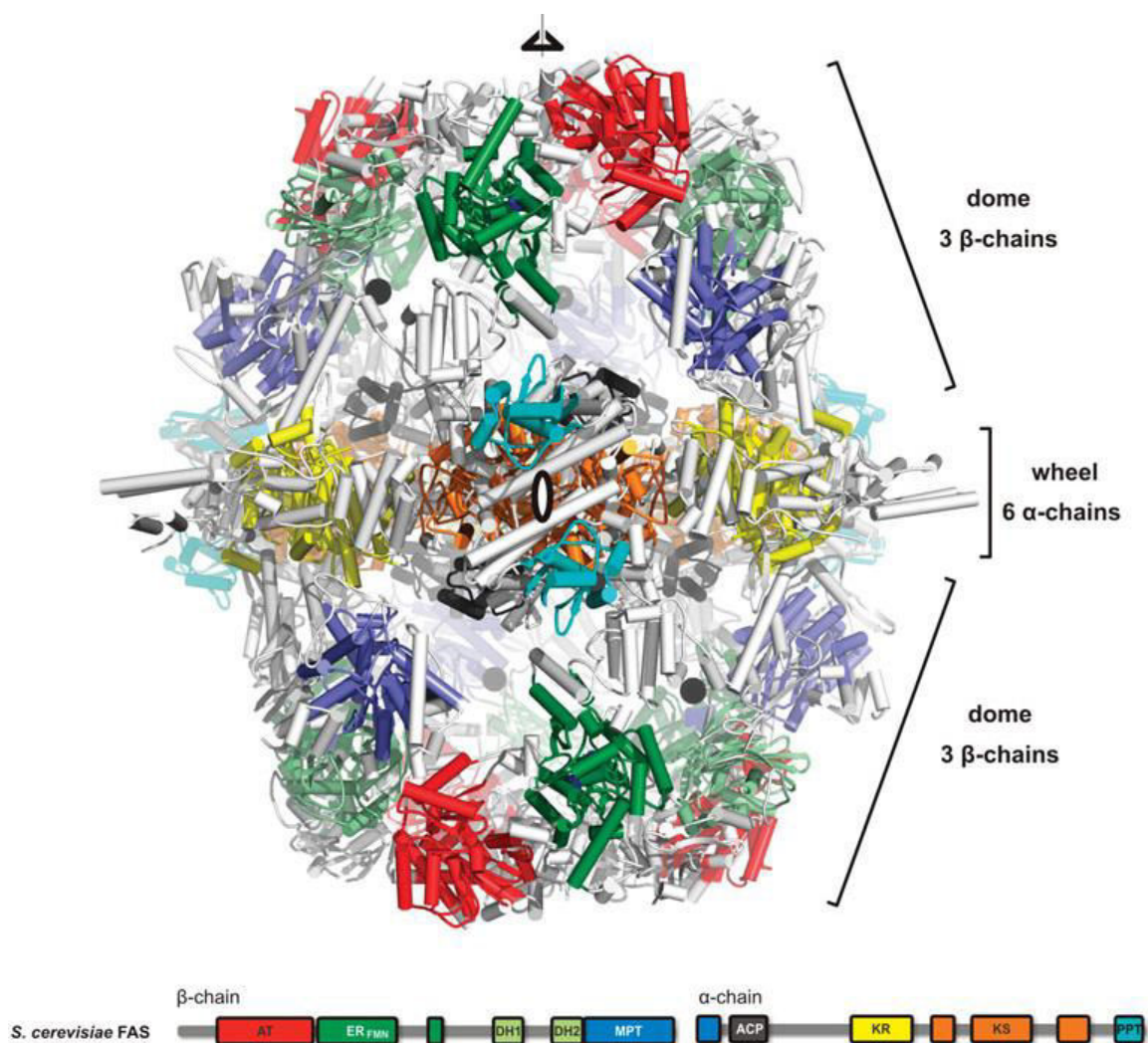


Figure 6 : Structure de la FAS de type I de *Saccharomyces cerevisiae* (Maier, *et al.*, 2010). Cette structure résolue à 3,1 Å de résolution est colorée par domaine suivant le schéma de la FAS situé en dessous. Les linkers entre les domaines sont indiqués en gris. Comme expliqué précédemment, la FAS s'organise en hétéro-dodécamère, un hexamère de la chaîne α et deux trimères de la chaîne β . La position de l'axe de symétrie d'ordre 3 et d'un des deux axes d'ordre 2 de la structure est indiquée

2. Les FAS animales, des synthèses homologues aux PKS

Au contraire des FAS fongiques, les FAS animales consistent en une chaîne polypeptidique de 270 kDa qui comprend tous les domaines nécessaires aux réactions de condensation et de réduction du C β . Malgré une organisation différente, les deux FAS partagent le même mécanisme d'action à quelques exceptions près. Une malonyl-acétyl-transférase (MAT) bi-fonctionnelle catalyse à la fois la réaction de transfert de l'unité de base (acétyl-CoA) et les réactions de transfert des unités d'extension (malonyl-CoA) sur l'ACP. En fin de synthèse, l'acide gras est libéré par un domaine supplémentaire par rapport aux FAS fongiques : le domaine thioestérase (TE). La dernière différence avec les FAS fongiques se situe au niveau du domaine ER, qui n'est pas FMN/NADPH dépendant mais nécessite uniquement le NADPH comme co-facteur (Smith, *et al.*, 2003) (Figure 7).

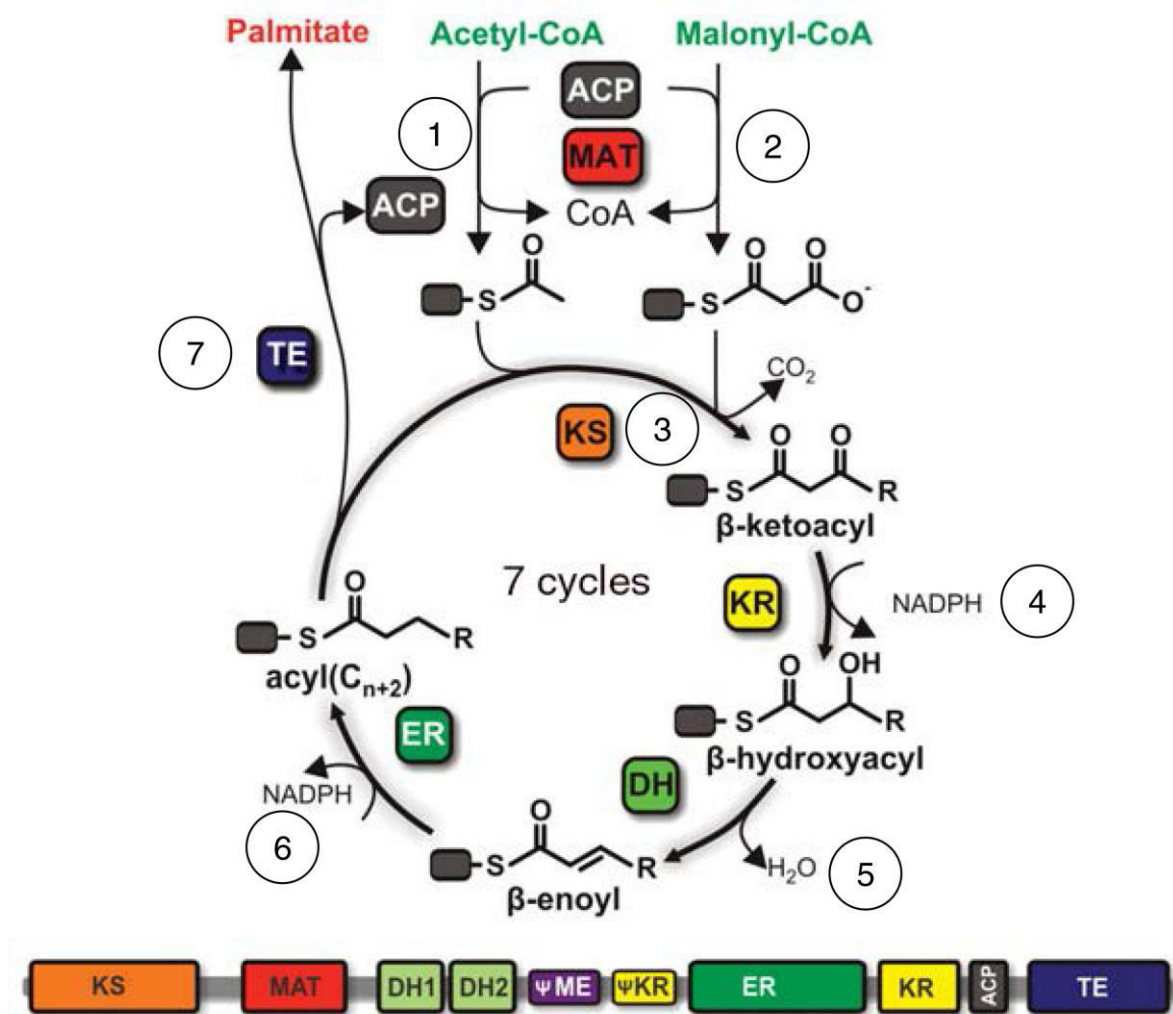


Figure 7 : Schéma réactionnel de la FAS fongique (adapté de Maier, et al., 2010).

Le MAT transfère l'unité de départ sur l'ACP (1) qui la transfère ensuite sur le KS. Le MAT charge alors l'unité d'extension sur l'ACP (2) et le KS catalyse la réaction de condensation des deux unités (3). Les réactions de réduction (4, 5 et 6) transforment la cétone en acyl saturé. Après 7 cycles, l'acide gras est libéré par un clivage catalysé par le domaine TE. Celui-ci libère du palmitate (ou stéarate) et non du palmitoyl-CoA comme c'était le cas pour les FAS fongiques. La composition en domaines est indiquée en dessous, en fonction des domaines découverts dans la structure de la FAS porcine.

La structure de la FAS porcine a été résolue en 2008 à 3.1 Å de résolution (Maier, et al., 2008). Dans cette structure, la FAS de mammifère (mFAS) forme un homodimère, les domaines KS, ER et DH étant dimériques. On observe une ségrégation entre les domaines responsables de la polymérisation (KS, MAT) et ceux assurant une activité de modification du groupement β-céto (KR, DH et ER). La connexion entre la partie condensation et celle faisant la modification est assurée par un linker de 20 acides aminés situé entre le linker post-MAT et le domaine DH. Ce linker étant non structuré, il permettrait un mouvement de la portion réductive par

rapport à la portion impliquée dans la condensation. Une étude par cryo-EM a en effet montré un mouvement de rotation des deux portions de la FAS l'une par rapport à l'autre (Brignole, *et al.*, 2009).

En plus des domaines catalytiques, d'autres domaines se situent aux extrémités de la portion réductrice de la mFAS. Un domaine pseudo kétoréductase (Ψ KR), analogue au domaine KR mais tronqué au niveau du Rossmann fold, assure une pseudo-homo-dimérisation du domaine KR. De même, un domaine structuralement proche des méthyltransférases à co-facteur *S*-adénosyl méthionine (SAM) est retrouvé au sein de la mFAS. Ce domaine appelé pseudo-méthyltransférase (Ψ ME) présente au sein du site actif un motif ExLxGxG au lieu du motif D/ExGxGxG conservé chez les enzymes SAM-dépendantes, ce qui empêche la fixation du co-facteur et donc rend ce domaine non-fonctionnel. Ces deux domaines non-fonctionnels ont sans doute un rôle uniquement structural, qui explique leur conservation au cours de l'évolution.

Un *linker* structuré d'une centaine de résidus sépare le KS et le MAT. Il est composé de deux hélices α et de trois brins β . Ce linker est en fait constitué de 70 résidus qui séparent le KS et le MAT et de 30 résidus séparant le MAT du DH. Cette dernière partie, moins structurée, vient se loger le long du KS et sert ensuite de linker entre les deux portions de la FAS. En revanche, la position de l'ACP et du TE n'ont pas pu être déterminé du fait de la flexibilité vraisemblablement trop importante de ces deux domaines (Figure 8).

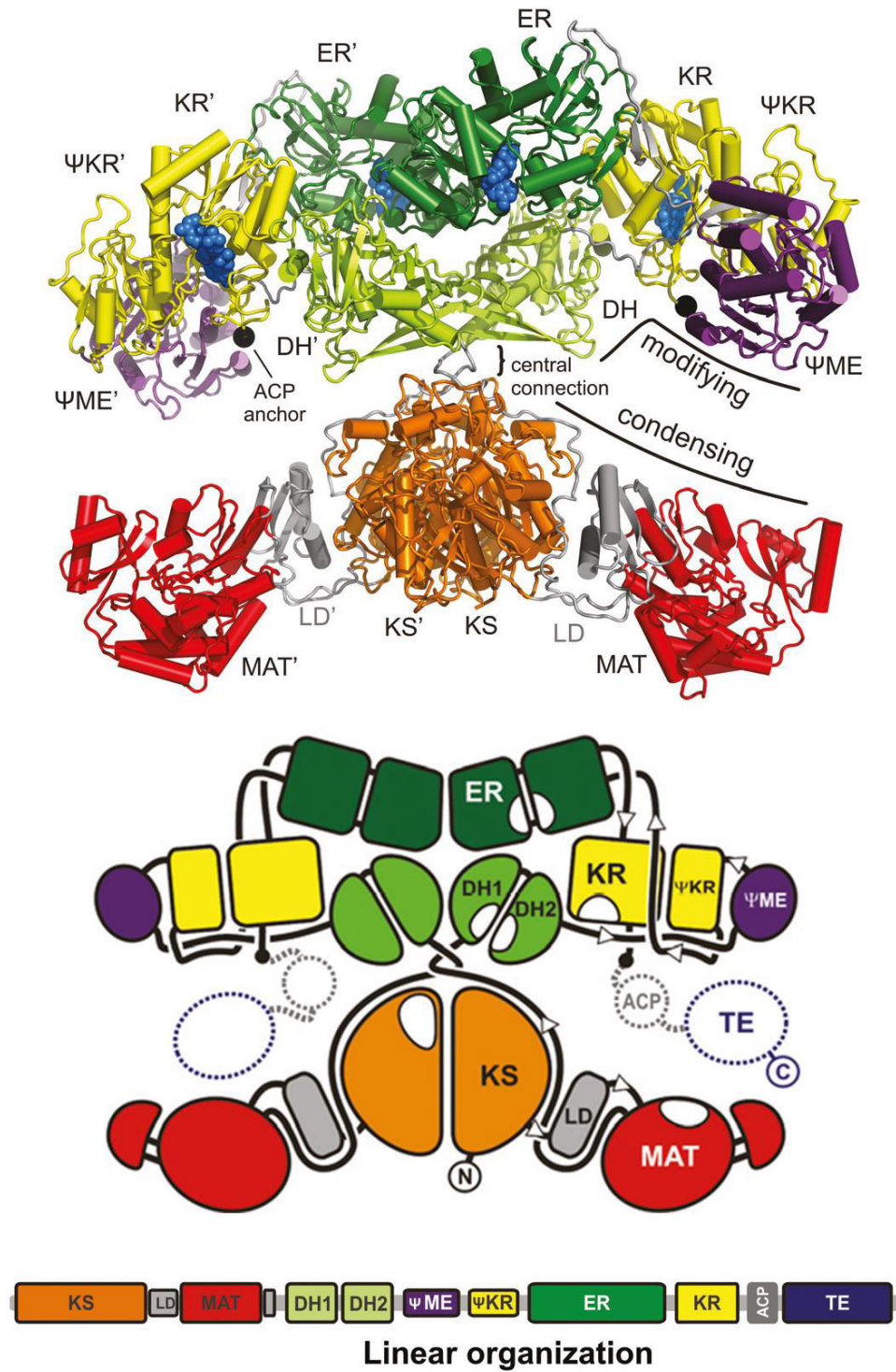


Figure 8 : Structure de la FAS animale résolue à 3,1 Å de résolution (adapté de Maier, et al., 2008)

Représentation cartoon de la mFAS, colorée par domaines. Les linkers séparant les domaines sont indiqués en gris. Les deux portions (condensation et modification) et le linker central séparant ces portions sont indiquées sur la structure, de même que la position du dernier résidu du linker entre KR et ACP (« ACP anchor »). On note que les domaines KS, DH et ER sont dimériques. En dessous, un schéma de la mFAS indique la position probable des domaines ACP et TE. LD signifie « linker domain » et indique le linker structuré entre le KS et le MAT.

Cette structure a représenté une énorme avancée pour les études sur la FAS humaine, qui est largement surexprimée dans les cellules cancéreuses, permettant ainsi le design d'inhibiteurs actuellement en essais cliniques (Flavin, *et al.*, 2010). De plus, cette structure est également un modèle pour les études des PKS de type I.

III. Les PKS

1. Organisation en cluster

Comme vu dans la partie I.2.B, les gènes codant les polykétide synthases s'organisent en cluster. Au sein de ces clusters se retrouvent typiquement tous les gènes codant les protéines servant à la synthèse même des polykétides, des protéines de maturation du polykétide une fois libéré par la PKS (enzymes post-PKS), les protéines de résistance au composé synthétisé, des éléments de régulation, etc...

Ainsi les clusters de gène codant les PKS peuvent occuper une part importante du génome. Pour exemple, la taille du cluster codant la synthèse d'érythromycine A est de 56 kb, alors que la totalité du génome de *Saccharopolyspora erythraea* compte 8 Mb, ce qui représente quasiment 1% du génome. D'ailleurs la totalité des clusters de gène responsable de la synthèse de polykétides par *S. erythraea* représente plus de 3% du génome de la bactérie (Oliynyk, *et al.*, 2007) (Figure 9).

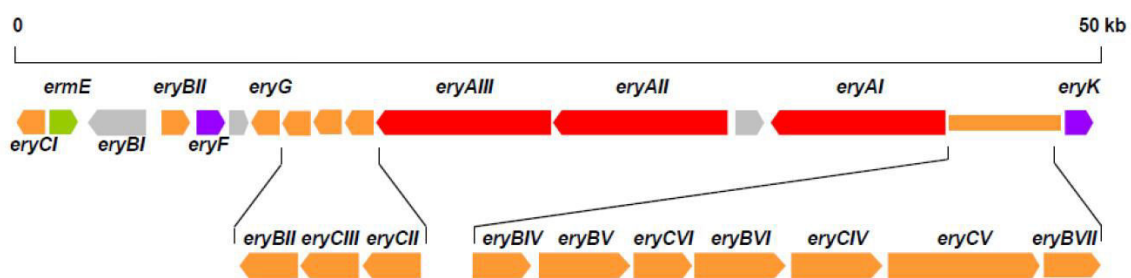


Figure 9 : Cluster de gènes responsable de la synthèse de l'érythromycine A (KJ Weissman).

Ce cluster de 56 kb code pour les trois chaînes polypeptidiques de la PKS (en rouge), les enzymes de modification post-PKS responsable de l'ajout d'unités glucidiques (en orange), des hydroxylases (en violet), et le gène conférant la résistance (en vert).

En fonction de leur organisation, tant au plan génique que protéique, et de leur mécanisme, les PKS ont été classées en trois types. En effet, lors de la découverte des clusters de gènes codant l'actinorhodine et celui codant l'érythromycine A, les chercheurs se sont aperçus que la première PKS était constituée de protéines discrètes (type II) alors que celle synthétisant l'érythromycine A est composée de

gigantesques multienzymes (type I). Cette classification a ainsi été établie en fonction des traits communs entre les PKS et les FAS (également divisées en deux types selon les mêmes critères, comme décrit précédemment). Cependant, il existe des PKS de type III composés d'une seule protéine responsable de la synthèse entière du polykétide.

2. Les PKS de type III

Ce type de PKS a été isolé pour la première fois chez *Petroselinum hortense*, une plante de la famille du persil. Elles sont responsables de la synthèse de nombreux métabolites secondaires chez les plantes tels que les résorcinols, les stilbènes, les biphényles, les benzophénones, les curcuminoïdes, etc... Nombre de ces composés sont des précurseurs de flavonoïdes ou de pigments. Bien qu'elles soient majoritairement décrites chez les plantes où elles sont ubiquitaires, les PKS de type III sont aussi présentes chez certains champignons ou certaines bactéries comme *S. griseus* (Kwon, *et al.*, 2001) .

Ce type de PKS, aussi couramment appelé stilbène synthase ou chalcone synthase, se différencie des autres PKS par la présence d'une unique protéine multifonctionnelle qui catalyse toutes les réactions nécessaires à l'élongation du polykétide. En effet, ces PKS ont la particularité de ne pas nécessiter d'ACP comme support de la chaîne polykétidique et des unités d'extension. Seul un KS dimérique multifonctionnelle va à la fois sélectionner les unités d'extension et catalyser les réactions de condensation et de cyclisation (Tropf, *et al.*, 1995, Preisig-Müller, *et al.*, 1997). Cette KS va donc agir de façon itérative afin de réaliser plusieurs cycles de condensation en utilisant des unités d'extension simplement fixés au coenzyme A et non à un ACP. Des enzymes de maturation post-PKS interviennent ensuite pour modifier le polykétide suite à sa libération (Figure 10).

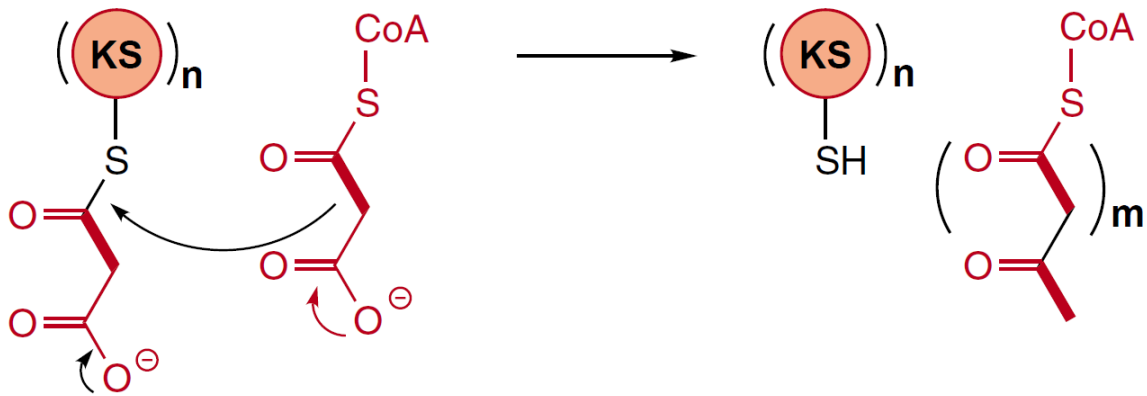


Figure 10 : Représentation schématique des PKS de type III (adapté de Shen, 2003).

Dans ce type de PKS, seule une KS dimérique agit de façon itérative pour catalyser la décarboxylation des unités d'extension liées au CoA et la condensation avec la chaîne polykétidique en élongation attachée à la cystéine du site actif.

3. Les PKS de type II :

Les PKS de type II, retrouvées chez certaines bactéries et champignons, synthétisent exclusivement des polykétides aromatiques. Par analogie avec les FAS de type II, ces PKS présentent des activités d'élongation de chaîne portées par des enzymes complètement dissociables qui interagissent en *trans* (Hertweck, *et al.*, 2007 ; McDaniel, *et al.*, 1995). Ces sous unités vont former un complexe composé d'au minimum trois partenaires.

Tout d'abord, une kétosynthase α ($KS\alpha$) va catalyser les réactions de condensation de façon itérative à partir d'unités d'extensions. Cette $KS\alpha$ va former un hétérodimère avec une kétosynthase β ($KS\beta$), aussi appelée « *Chain Length Factor* (CLF) ». Cette $KS\beta$ est dépourvue de cystéine catalytique et permet de contrôler la longueur de la chaîne polykétidique soit en orientant les réactions vers un nouveau cycle de polymérisation ou en déclenchant la libération et la cyclisation du polykétide (Keatinge-Clay, *et al.*, 2004). Il est également démontré que la $KS\beta$ initie la synthèse en catalysant la décarboxylation du malonyl-CoA en acétyl-CoA qui va servir d'unité de départ à la polymérisation (Bisang, *et al.*, 1999). De plus, l'hétérodimère $KS\alpha/KS\beta$ serait responsable du contrôle de la régiospécificité de la première réaction de cyclisation au cours de la synthèse du polykétide (Hertweck, *et al.*, 2007). Enfin, un ACP va porter les unités d'extension et la chaîne en élongation au cours de la synthèse (Figure 11).

En revanche, la façon dont les unités d'extension sont transférées sur l'ACP demeure encore incertaine. Il a été proposé que l'ACP puisse avoir une activité automalonylase, et serait donc capable de sélectionner une unité malonate à partir de fortes concentration en malonyl-CoA et ce, en absence d'AT (Hitchman, *et al.*, 1998) (Arthur, *et al.*, 2005). Il est également proposé que ce soit une malonyl-CoA:ACP transférase (MCAT) absente du cluster de gène codant la PKS mais provenant de la FAS (de type II), qui serait responsable du transfert du malonyle sur l'ACP. Il a également été supposé que, du fait de la présence du motif GHSxG propre aux AT, les KS pourraient également présenter une activité acyltransférase (Fernandez-Moreno, *et al.*, 1992).

Des protéines de modification optionnelles interviennent à certaines étapes de l'élongation. Des kétéroréductases réduisent la fonction cétone de l'unité d'extension ajoutée en fonction alcool, des cyclases jouent le rôle de « chaperonnes moléculaires » pour favoriser des réactions de cyclisation spontanée, et des aromatasés catalysent des réactions de déshydratation au niveau du groupement alcool (donc après action d'une kétéroréductase) afin de former un cycle aromatique.

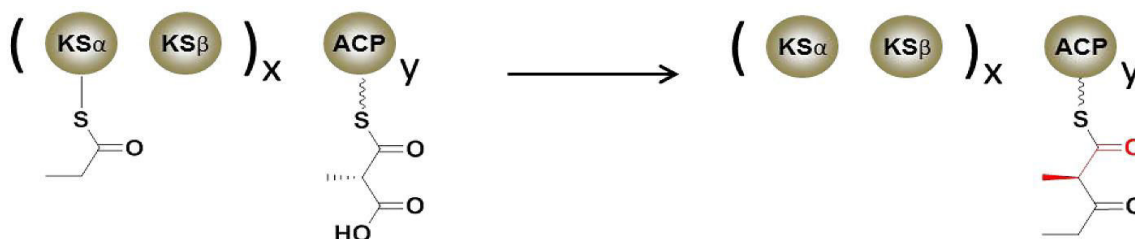


Figure 11 : Représentation schématique des PKS de type II (tiré de Shen, 2003).

Dans ce type de PKS, un hétérodimère composé de KS α et de KS β (Chain length factor) interagit avec un ACP en *trans* pour catalyser des réactions de condensation et ce, de manière itérative. L'unité acyl ajoutée à chaque cycle d'élongation est indiquée en rouge. Des enzymes de modification optionnelles interviennent également au cours du processus d'élongation.

4. Les PKS de type I

Ce type de PKS présente beaucoup d'homologie avec les FAS de type I animale décrite précédemment, tant au niveau séquence, qu'au niveau mécanistique. La principale différence est qu'au contraire des FAS, les PKS de type I sont des protéines multi-modulaires. Ainsi chaque module catalyse l'incorporation d'une unité acyl-CoA et donc l'élongation de la chaîne polykétidique. Chaque module comporte lui-même

plusieurs domaines responsables d'une activité enzymatique propre. Trois domaines sont essentiels : les domaines AT, ACP et KS. Trois domaines supplémentaires, les domaines réducteurs KR, DH et ER, sont optionnels au sein des modules (Fischbach, et Walsh, 2006) (Figure 12).

Le fait que ces domaines soient optionnels permet la création d'une diversité de structures des polykétides synthétisés, ces modifications étant conservés dans la forme mature du polykétide. Un domaine TE, présent au sein du dernier module, a pour fonction de libérer la chaîne polykétidique et, dans la majorité des cas, d'entraîner sa cyclisation. Ainsi, une même protéine va regrouper un ou plusieurs module(s) responsable(s) de l'incorporation d'une unité d'extension à la chaîne, agissant alors comme une usine d'assemblage. Par conséquent, les PKS de type I sont des protéines de très haut poids moléculaire, allant jusqu'à 1,8 MDa pour une des PKS formant la mycolactone (Stinear, *et al.*, 2004).

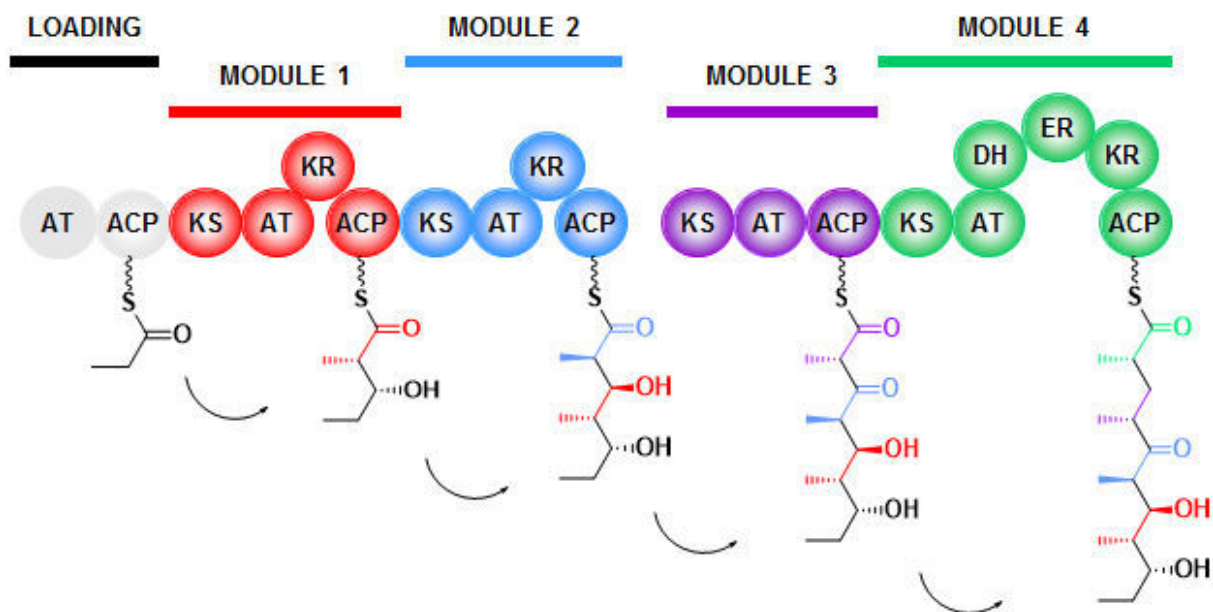


Figure 12 : Représentation schématique des PKS de type I (figure réalisée par KJ. Weissman).

Les PKS de type I adoptent une organisation modulaire. Dans le cas présent, la PKS regroupe deux protéines, l'une comprenant le module de chargement et les deux premiers modules, et la deuxième qui comprend les modules 3 et 4. Chacun de ces modules est composé d'un certain nombre de domaines. La chaîne en élévation est transférée d'un module au suivant après chaque cycle d'élévation

A) Les PKS de type I itératives

Il existe des PKS de type I qui, comme les FAS de type I animale, ne sont composées que d'un module qui agit de façon itératif. Ce type de PKS permet notamment la synthèse d'acides gras polyinsaturés (PUFA) qui sont des constituants des membranes cellulaires de certaines bactéries marines (Metz, *et al.*, 2010). Elles permettent également la synthèse de polykétides aromatiques chez les bactéries ou de composés de la famille des ènediynes (des composés organiques contenant deux triples liaisons et une double liaison) (Ahlert, *et al.*, 2002).

Chez les champignons, les PKS de type I itérative sont classés en trois groupes en fonction des activités réductrices (KR, DH et ER) intervenant durant l'élongation. Ainsi on trouve les PKS non réductrice (NR), partiellement réductrices (PR) et hautement réductrices (HR) (Cox, 2007).

B) Les PKS de type I modulaires

Au contraire des PKS de type I itératives, ce type de PKS est composé de plusieurs modules qui vont être chacun responsable d'un cycle d'élongation. Chaque module va être composé d'un set différent de domaines pour chaque cycle d'élongation. Ainsi chaque PKS peut incorporer un nombre différent de modules eux même composés d'un panel variable d'activités enzymatiques. Parmi ces PKS modulaires on distingue deux types de PKS différentes.

a. Les PKS de type I modulaires de type *cis*-AT

Chez les PKS de type *cis*-AT, les domaines servant à la condensation (KS, ACP, AT), ceux catalysant les réactions de modification durant la synthèse du polykétide (KR, DH, ER) et même le domaine de libération du polykétide en fin de synthèse (TE) sont situés au sein d'une même chaîne polypeptidique.

Depuis sa découverte en 1990, la PKS synthétisant l'érythromycine A (DEBS pour désoxyérythronolide B synthase) est devenue le système modèle pour les PKS de type *cis*-AT. Ce fût en effet la première PKS de type I découverte (Cortés, *et al.*, 1990 ; Donadio, *et al.*, 1991), et de plus, elle constitue un modèle parfait pour l'étude de l'ensemble des PKS de type *cis*-AT. Cette PKS est constituée de 3 chaînes polypeptidiques : DEBS 1 qui porte le module de chargement, nécessaire pour recruter

l'unité de départ et la transfère sur le premier module, ainsi que les deux premiers modules, DEBS 2 qui porte les modules 3 et 4 et DEBS 3 composée des module 5 et 6 comportant ainsi l'activité TE nécessaire à la libération du polykétide.

La synthèse débute par le recrutement d'un propionyl-CoA par le domaine AT du module de chargement qui le transfère alors sur l'ACP voisin. L'AT du module 1 sélectionne un nouveau methylmalonyl-CoA comme unité d'extension et le transfère sur l'ACP du module 1. Dès lors, le KS catalyse la condensation des deux unités sélectionnées afin de former un dikétide. Le domaine KR, présent au sein du module 1, catalyse la réduction du groupement β -céto en hydroxyle. Dès lors, la chaîne en extension est transférée au niveau du KS du module 2 et un nouveau cycle d'extension peut avoir lieu. Il y a ainsi 6 cycles d'extension et en fin de synthèse, le domaine TE présent en C-terminale du module 6 catalyse la libération du polykétide et sa cyclisation afin de former le 6-désoxyérythronolide B. Cet intermédiaire est ensuite mûri par des enzymes post-PKS qui vont, entre autres, ajouter deux unités glycosidiques (Figure 13).

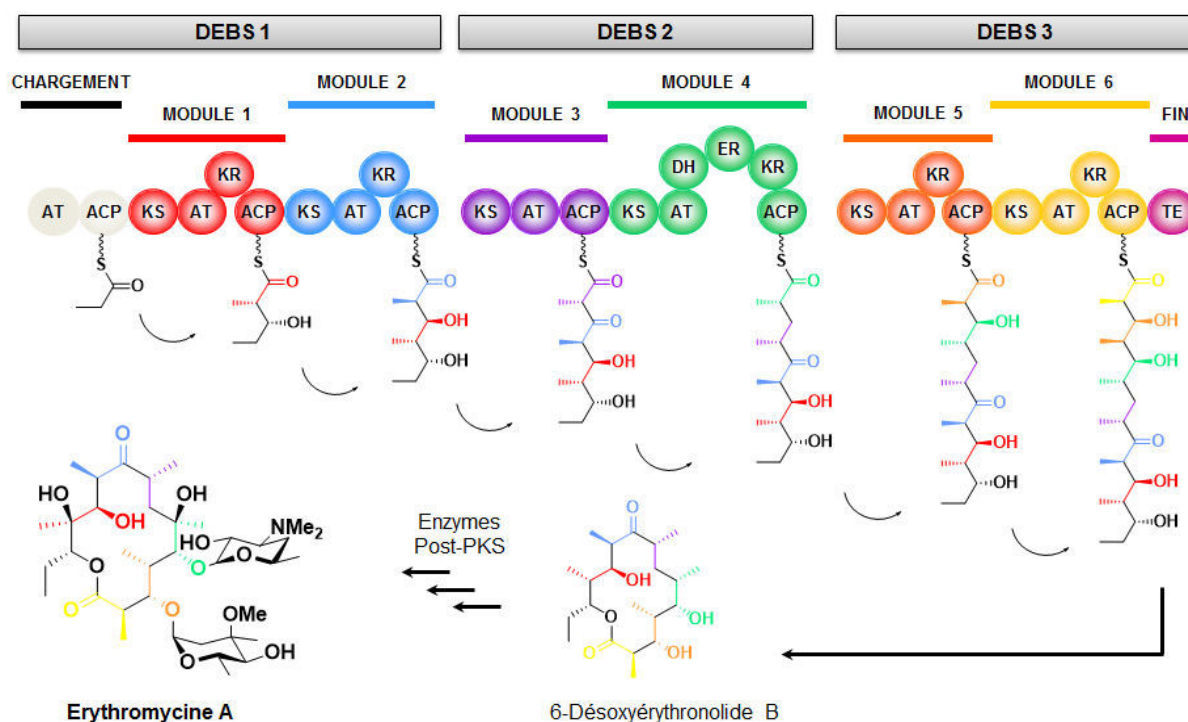


Figure 13 : Représentation schématique de la PKS synthétisant l'érythromycine A (figure réalisée par KJ Weissman).

La PKS est composée de 3 sous-unités, DEBS 1, 2 et 3. Chaque sous-unité comprend plusieurs modules, chacun représenté par une couleur différente. L'unité d'extension ajoutée au sein du polykétide en cours d'élongation est représentée de la même couleur que le module catalysant cet ajout. Un intermédiaire, le 6-désoxyérythronolide B est libéré par le domaine TE, puis subi une maturation par des enzymes post-PKS pour former l'érythromycine A.

Dans ce type de PKS, il existe une relation directe entre la séquence primaire de la PKS et la nature du composé synthétisé. Ainsi, chaque module agit dans l'ordre selon lequel il est situé au sein de la séquence polypeptidique, l'enchaînement en domaines des modules dictant alors la structure du polykétide synthétisé. Cette relation est décrite sous le terme de règle de colinéarité et est propre aux PKS de type *cis*-AT. Ceci a été démontré par des expériences d'inactivation d'une KR (Donadio, *et al.*, 1991) ou d'une ER (Donadio, *et al.*, 1993) dans les modules 5 et 4 respectivement de la PKS responsable de la synthèse de l'érythromycine. Ainsi, des dérivés ont été obtenus avec absence de modifications aux positions escomptées. Dans ce cas, la modification d'un domaine ou d'un module entraîne un changement direct et prévisible de la structure du polykétide à la position voulue, d'où l'engouement que l'étude de ces enzymes suscite en vue d'approches d'ingénierie. Ce constat explique pourquoi les PKS de type *cis*-AT sont les plus étudiées à l'heure actuelle.

Cette relation a également permis de développer des outils de bio-informatique afin de prédire la structure de polykétides synthétisés par des espèces dont le génome a été séquencé (Bachmann, et Ravel, 2009).

b. Les PKS de type I modulaires de type *trans*-AT

Pour certaines PKS de type I, aucun domaine AT n'a été identifié dans la séquence primaire des modules. Bien que les mécanismes de synthèse soient identiques aux PKS de type *cis*, il a été découvert que l'acylation des domaines ACP nécessite une protéine isolée, également appelée acyltransférane, qui interagit en *trans* avec les ACP de plusieurs modules pour apporter chaque fois la même unité d'extension (Piel, 2002). Ce type de PKS fut découvert pour la première fois dans le génome séquencé de *Bacillus subtilis* en 1993 (Scotti, *et al.*, 1993). Cependant, les chercheurs ne sont pas parvenus à identifier un polykétide produit par cette souche et le séquençage comportait de multiples erreurs (Scotti, *et al.*, 1993). La première PKS *trans*-AT rattachée à la production d'un polykétide fut mise en évidence chez *Paederus fuscipes*, une espèce de scarabée, en 2002. Cette PKS responsable de la synthèse de pédérine, un anti-cancéreux, ne disposait pas d'AT au sein de la séquence des modules et il fut supposé que c'est une AT discrète, dont le gène est présent au sein du cluster codant la PKS, qui acyle les ACP (Piel, 2002). Ce type d'acylation fut confirmé *in vitro* pour la PKS responsable de la synthèse de leinamycine (Cheng, *et al.*, 2003).

Les PKS de type *trans*-AT présentent de nombreuses particularités par rapport aux PKS de type *cis*-AT. Comme dit précédemment, la protéine AT sélectionne la même unité d'extension à chaque cycle de condensation. La Nature a donc trouvé d'autres moyens pour générer de la diversité. En effet, on trouve parmi ces PKS un domaine de modification optionnelle supplémentaire, le domaine méthyltransférase (MT). Ce domaine catalyse l'ajout d'un groupement méthyle en position α grâce à l'utilisation d'un co-facteur commun aux méthyltransférases, la SAM.

Une autre particularité est la présence de modules de synthétases peptidiques non ribosomiques (NRPS), même si de rares PKS de type *cis*-AT en possèdent également. Ces synthétases, analogues aux PKS, permettent d'incorporer des acides aminés au lieu d'unités acyles afin de former des peptides non ribosomiques (NRP). Parmi ces NRP on retrouve également des composés de forte valeur ajoutée tel que la pénicilline G, le premier antibiotique découvert par Flemming, la vancomycine (un antibiotique utilisé contre des infections sévères par des bactéries Gram +), et bien d'autres composés dont l'activité pharmacologique se révèle être intéressante.

Les NRPS, comme les PKS, sont composés de plusieurs modules catalysant chacun l'incorporation d'un acide aminé au sein du peptide en extension. Trois domaines sont nécessaires : le domaine d'adénylation (A), analogue au domaine AT, qui active l'acide aminé par utilisation concomitante d'ATP. Il adényle l'acide aminé par transfert d'AMP au niveau du groupement carboxyle afin de former un aminoacyl-adénylate réactif. Puis, le domaine A catalyse le transfert de l'acide-aminé adénylé sur le domaine peptidyl carrier protein (PCP), analogue de l'ACP, qui va servir de support aux unités d'extension et au peptide en cours d'élongation. Et enfin, un domaine de condensation (C), analogue au domaine KS, va catalyser les réactions de condensation entre les acides aminés liés aux PCP pour former des liaisons peptidiques (Figure 14).

Des domaines de modification optionnels permettent soit l'ajout d'un groupement méthyle par l'intermédiaire du domaine méthyltransférase (MT) (Burmester, *et al.*, 1995), ou bien un domaine épimérase (E), transformant ainsi l'acide-aminé L en acide aminé D (Stachelhaus, et Walsh, 2000), ou encore des domaines de hétérocyclisation (Cy ou HC) remplaçant le domaine de condensation au sein du module et permettant la condensation et la formation subséquente d'un cycle thiazole lors de l'incorporation

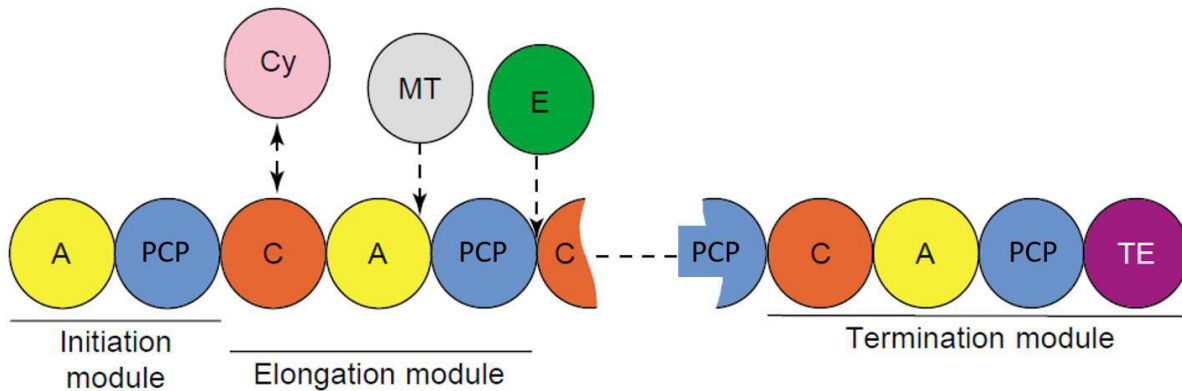


Figure 15 : Représentation schématique synthétisant la composition en domaines des NRPS (Challis, et Naismith, 2004).

Un module d'initiation composé des domaines d'adénylation (A) et *peptidyl carrier protein* (PCP) permet de charger le premier acide aminé et ensuite plusieurs modules effectuent l'élongation du peptide grâce à la formation de liaison peptidique catalysée par le domaine de condensation (C). Des domaines optionnels peuvent également intervenir comme les domaines de hétérocyclisation (Cy ou HC), d'épimérisation (E) et de méthylation (MT). Au sein du dernier module, un domaine thioestérase (TE) permet la libération du peptide.

Ainsi, la présence de module NRPS au sein des PKS de type *trans*-AT permet l'incorporation d'un acide aminé au sein du polykétide. Ces hybrides PKS-NRPS permettent donc la génération d'une grande diversité de structures.

De plus, au sein des PKS de type *trans*-AT, il est assez fréquent de trouver des modules scindés sur deux chaînes polypeptidiques qui se complètent en *trans*. On retrouve aussi dans la séquence primaire de ces PKS, des domaines voire des modules entiers ne présentant pas d'activité et dont la fonction doit être uniquement d'ordre structurale pour expliquer leur conservation au cours de l'évolution. On trouve également chez certaines PKS des modules qui agissent de façon itérative pour catalyser plusieurs cycles d'élongation (Piel, 2010).

Par conséquent il est très difficile de prédire la structure du polykétide synthétisé simplement en regardant la structure primaire de la PKS. En effet, du fait de ces particularités, la règle de colinéarité n'est pas valable pour les PKS de type *trans*-AT contrairement à celles de type *cis*. Pour pallier à ce manque d'information, une méthode phylogénétique pour prédire la structure du polykétide synthétisé a été développée (Nguyen, *et al.*, 2008). Celle-ci s'appuie sur l'étude évolutive des domaines KS. Ainsi, par l'analyse de séquences des KS, ces derniers sont rangés en différents clades, chaque clade acceptant un substrat particulier. Il est alors possible

de prédire les groupements chimiques constituant le polykétide au stade de synthèse où le module comportant le KS intervient, et ainsi de prédire sa structure. Cette analyse phylogénétique a également montré que les PKS *trans*-AT semblent avoir évolué d'une façon différente des PKS *cis*-AT (Nguyen, *et al.*, 2008). Alors que ces dernières auraient évolué principalement par duplication de gènes, les *trans*-AT sembleraient issues d'échanges horizontaux de gènes entre bactéries, formant ainsi une mosaïque de gènes (Piel, 2010). Ainsi, les connaissances dont nous disposons pour le type *cis*-AT ne sont pas toujours directement applicables pour le type *trans*-AT ; l'étude indépendante des deux systèmes est donc nécessaire.

La PKS synthétisant la virginiamycine M, un antibiotique, est un bon exemple de PKS *trans*-AT. Son fonctionnement sera développé plus en détail dans le chapitre IX.1 page 75 (Figure 16).

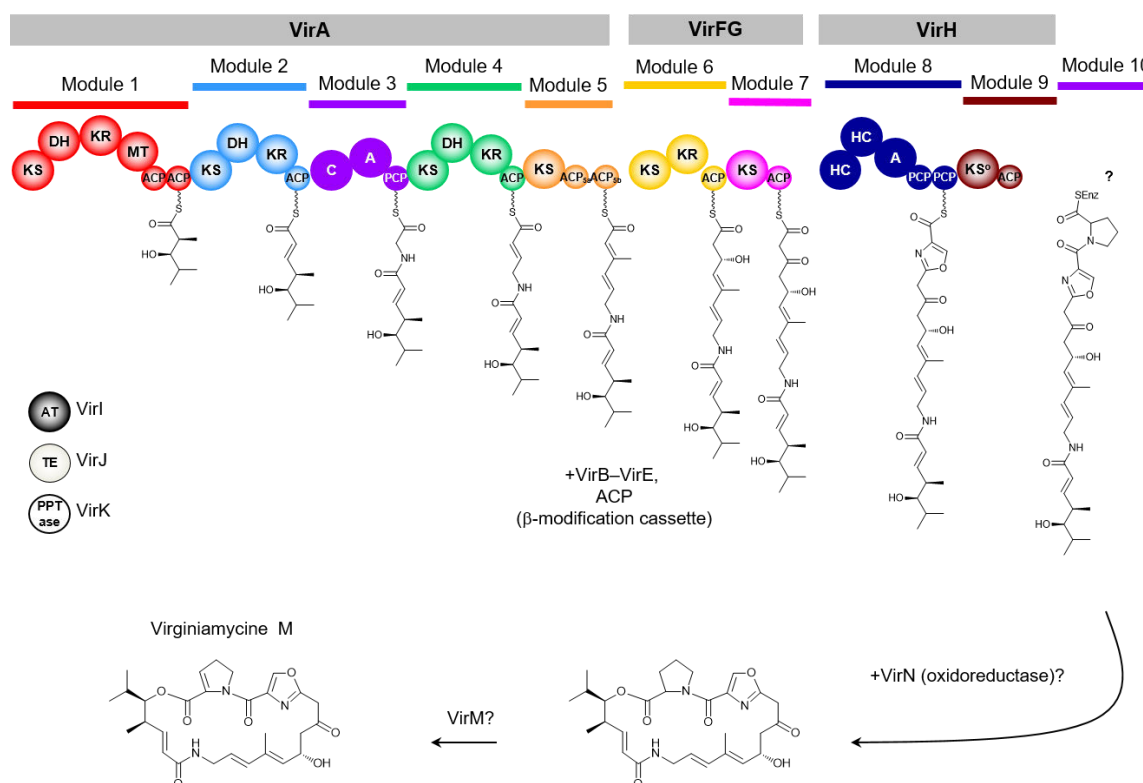


Figure 16 : Exemple de PKS de type *trans*-AT avec la PKS synthétisant la virginiamycine M chez *S. virginiae*.

Cette PKS présente une protéine acyltransférase discrète (VirI), un domaine MT au sein du module 1 qui permet l'ajout d'un méthyle sur le premier malonyl incorporé au sein du polykétide, trois modules NRPS (modules 3, 8 et 10), un module inactif (module 9) qui comporte une KS (appelée KS⁰) dont la cystéine catalytique est absente, et un TEII discret (VirJ). HC : domaine d'hétérocyclisation. Un module de chargement a également été trouvé dans le cluster de gènes homologue chez *S. pristinaspiraelis*.

IV. L'attractivité des PKS de type I pour la biologie synthétique

Du fait de leur fort potentiel pharmacologique, la découverte de nouveaux polykétides suscite un énorme intérêt. Cette découverte passe par trois voies possibles : 1) la recherche de nouvelles molécules au sein même de la biosphère, une approche facilitée par le développement du *genome-mining* (Bachmann, *et al.*, 2014 ; Monciardini, *et al.*, 2014). En effet, par le passé, la découverte de nouveaux composés se faisait généralement au fruit du hasard, guidé par la bioactivité de certains extraits de culture en fermentation ou bien par un crible chimique des métabolites avec des propriétés chimiques particulières (Bachmann, *et al.*, 2014). L'essor du séquençage et le développement du *genome-mining* a ainsi grandement facilité la découverte de polykétides au sein de la biosphère.

2) Une autre façon de générer de nouveaux polykétides consiste en la synthèse héli-chimique de composés existant naturellement. En effet, étant donné la complexité structurale des polykétides, leur synthèse chimique complète est difficile et se révèle être polluante. Ainsi, l'héli-synthèse chimique consiste en l'ajout de groupement fonctionnel ou la modification d'un groupement préexistant sur un composé naturellement synthétisé, et permet par exemple d'éviter une résistance à un antibiotique ou de rendre plus affiné un composé envers sa cible (Nilius, et Ma, 2002). Cette approche rationnelle a déjà abouti à la synthèse de plusieurs composés, comme l'EM900, un dérivé héli-synthétique de l'érythromycine A, qui sert comme anti-inflammatoire (Sugawara, *et al.*, 2011), ou encore le Zithromax et la Biaxine décrits précédemment (Peltier, 1992).

3) Enfin la dernière stratégie, complémentaire aux deux précédentes, consiste en l'ingénierie des PKS.

1. Les PKS de type *cis*-AT, un modèle pour l'approche d'ingénierie

Du fait de leur modularité, les PKS de type I constituent un bon modèle pour la biologie synthétique. Cependant, jusqu'à maintenant, ces approches ont surtout été menées sur les PKS de type *cis*-AT. Ceci trouve une raison historique, puisque les PKS de type *cis* ont été les premières découvertes, mais également pratique en raison de la simplicité d'étude conférée par la règle de colinéarité. De ce fait, le seul prérequis pour effectuer une modification d'un groupement ou d'une position particulière du polykétide est d'identifier les domaines ou les enzymes de maturation post-PKS qui en sont responsables

Plusieurs modifications peuvent être apportées à la voie de synthèse d'un polykétide (Weissman, 2016) (Figure 17). Le plus intuitif, et ce qui semble être le plus probant, est de manipuler les domaines PKS de façon individuelle. Une des stratégies consiste en la mutation spécifique de certains résidus au sein du site actif de façon à altérer la spécificité de substrat, la stéréochimie de la réaction catalysée ou bien d'inactiver le domaine. Il est également possible d'échanger un domaine d'un module par un domaine provenant d'un autre module ou d'une autre PKS pour aboutir aux mêmes résultats que précédemment. Par exemple, l'échange d'une AT présentant une spécificité pour le malonyl-CoA par une AT spécifique pour le méthylmalonyl-CoA permettrait l'ajout d'un groupement méthyl au squelette polykétidique.

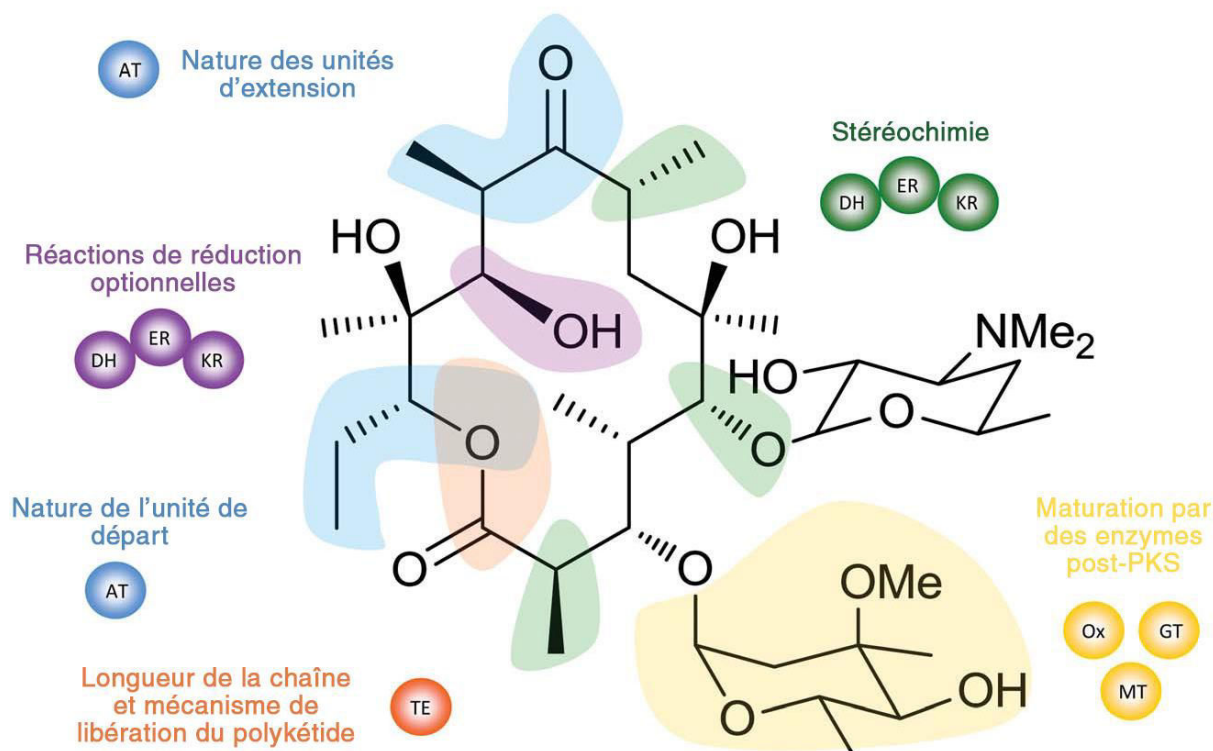


Figure 17 : Possibilités de modification des PKS pour la génération de nouveaux polykétides, illustré pour l'érythromycine (adapté de Weissman, 2016).

Le type de modifications de la PKS et les changements structuraux apportés au polykétide sont corrélés suivant un même code couleur. Il est à noter que toutes ces modifications sont théoriquement cumulables.

Plusieurs manipulations génétiques de PKS ont été réalisées à ce jour aboutissant à plus ou moins de succès. De nombreuses expériences d'ingénierie ont ainsi été menées et seules les plus notables seront présentées ici.

Tout d'abord, des mutations au niveau du site de fixation du NADPH des domaines KR et ER de DEBS ont permis la synthèse de dérivés de l'érythromycine A avec les modifications aux positions attendues (Donadio, *et al.*, 1991 ; Donadio, *et al.*, 1993). En revanche, une tentative pour muter un domaine DH de DEBS s'est révélée infructueuse. Bien que peu invasive, puisqu'il s'agit uniquement d'une mutation de l'histidine catalytique du domaine DH, aucun produit complet ne put être obtenu, probablement à cause de la spécificité des domaines composant le module suivant la modification (Bevitt, *et al.*, 1993).

Une étude plus récente a consisté à muter les résidus catalytiques de chaque domaine réducteur des modules 2 à 8 de la PKS synthétisant la monensine, soit 18 domaines réducteurs. Sur les 32 produits attendus, 22 ont pu être obtenus, mais avec

chaque fois un rendement au minimum 7 fois inférieur à celui de la PKS sauvage (Kushnir, *et al.*, 2012).

D'autres études ont eu pour but d'introduire des mutations au sein du domaine AT afin d'altérer la spécificité de ce dernier, et donc permettre la sélection d'unités d'extension différentes. Ceci a été réalisé avec succès pour l'AT₅ de la PKS de la monensine, pour laquelle en mutant une valine du site actif, ils sont parvenus à modifier la spécificité de l'AT, lui permettant d'incorporer préférentiellement du 2-propargylmalonate en tant qu'unité d'extension (Bravo-Rodriguez, *et al.*, 2014). Cependant les rendements obtenus suite à cette modification demeurent toujours inférieurs à ceux obtenus avec la PKS sauvage. Enfin, d'autres études ont eu pour objectif d'altérer la stéréospécificité des domaines ER et KR que ce soit par mutagenèse dirigée ou par échange de domaines (Kwan, *et al.*, 2008 ; Annaval, *et al.*, 2015).

Des tentatives pour insérer des modules entiers provenant d'autres PKS ont également été réalisées, permettant la création de PKS chimères. L'échange du module de chargement permet de modifier l'unité de départ, comme ce fut le cas chez DEBS où le module de départ a été échangé avec celui de la PKS synthétisant l'avermectine (Marsden, *et al.*, 1998).

Des échanges de module d'extension ont aussi été réalisés chez DEBS permettant d'obtenir les produits escomptés, à condition de conserver les *linkers* reliant les modules naturellement présents (Gokhale, *et al.*, 1999). Une autre tentative pour insérer un module provenant de la PKS synthétisant la rapamycine entre les modules 1 et 2 de DEBS, dans le but de rajouter deux carbones supplémentaires au polykétide, s'est révélée infructueuse. En effet, la chaîne en élongation était transférée directement d'un ACP à l'autre, court-circuitant ainsi le module introduit (Rowe, *et al.*, 2001 ; Thomas, *et al.*, 2002). Ainsi, bien que le produit attendu ait tout de même été obtenu de façon très minoritaire, le produit majoritaire correspond à un dikétide formant une lactone, résultant du saut du module inséré (Figure 18).

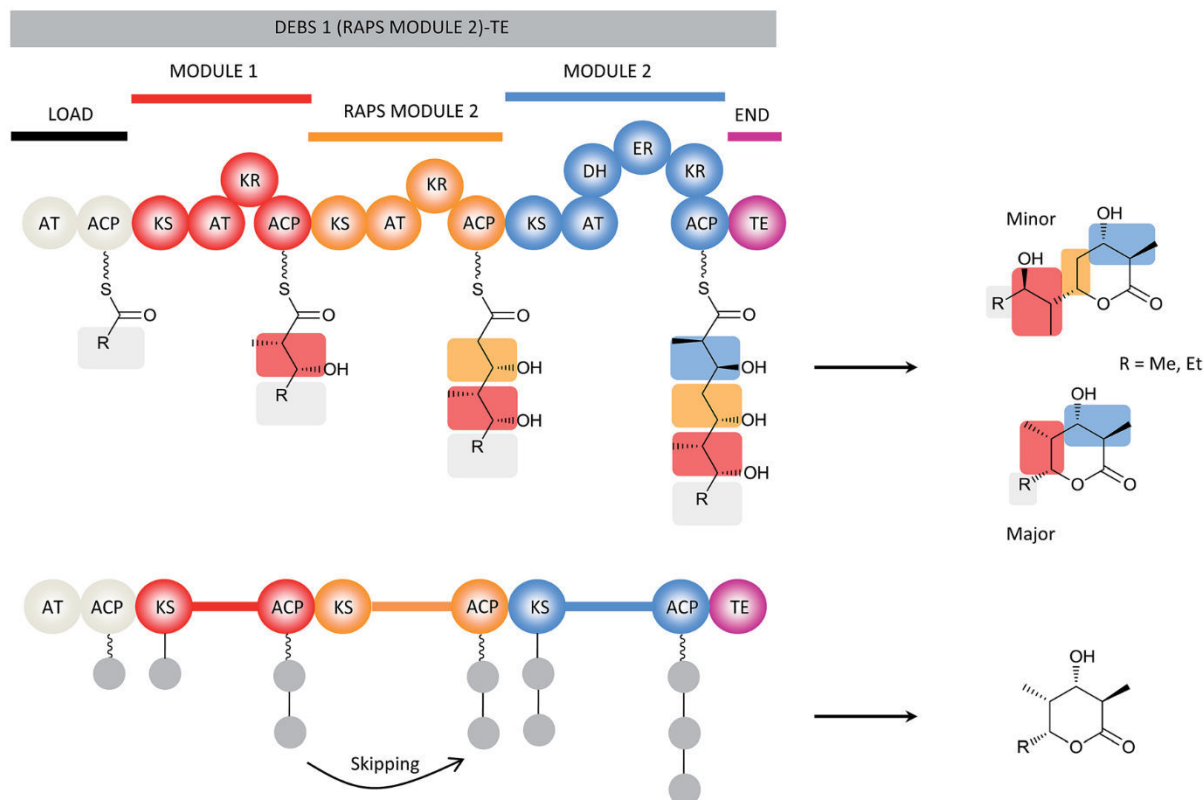


Figure 18 : Tentative d'ajout d'un module à la PKS synthétisant l'érythromycine (tiré de Weissman, 2016).

Le module 2 de la PKS synthétisant la rapamycine (RAPS) a été ajouté à DEBS 1-TE, une construction permettant d'obtenir une lactone à partir des deux premiers modules de la PKS synthétisant l'érythromycine auxquelles est fusionné le domaine TE. Les couleurs indiquent quels portions du polykétide chaque module est censé ajouter. Le composé attendu est synthétisé de façon minoritaire par rapport au trikétide. L'explication de la synthèse majoritaire de cette lactone est fournie dans le schéma en dessous dans lequel on peut voir que la chaîne passe d'un ACP à l'ACP du module inséré sans que celui-ci puisse catalyser un cycle d'élongation.

Une autre tentative d'échange de modules a été entreprise sur la PKS de la néoauréothine. Celle-ci consistait à conserver l'interface ACP–KS entre les modules 1 et 2. Le reste des domaines du deuxième module a alors été remplacé par le module 3 ou le module 4 (Figure 19). La construction dans laquelle les modules 2 et 3 ont été délévés n'a donné aucun produit. En revanche celle où seul le module 2 est supprimé a donné un nouveau polykétide plus court appelé homoauréothine avec un très faible rendement (Traitcheva, *et al.*, 2007).

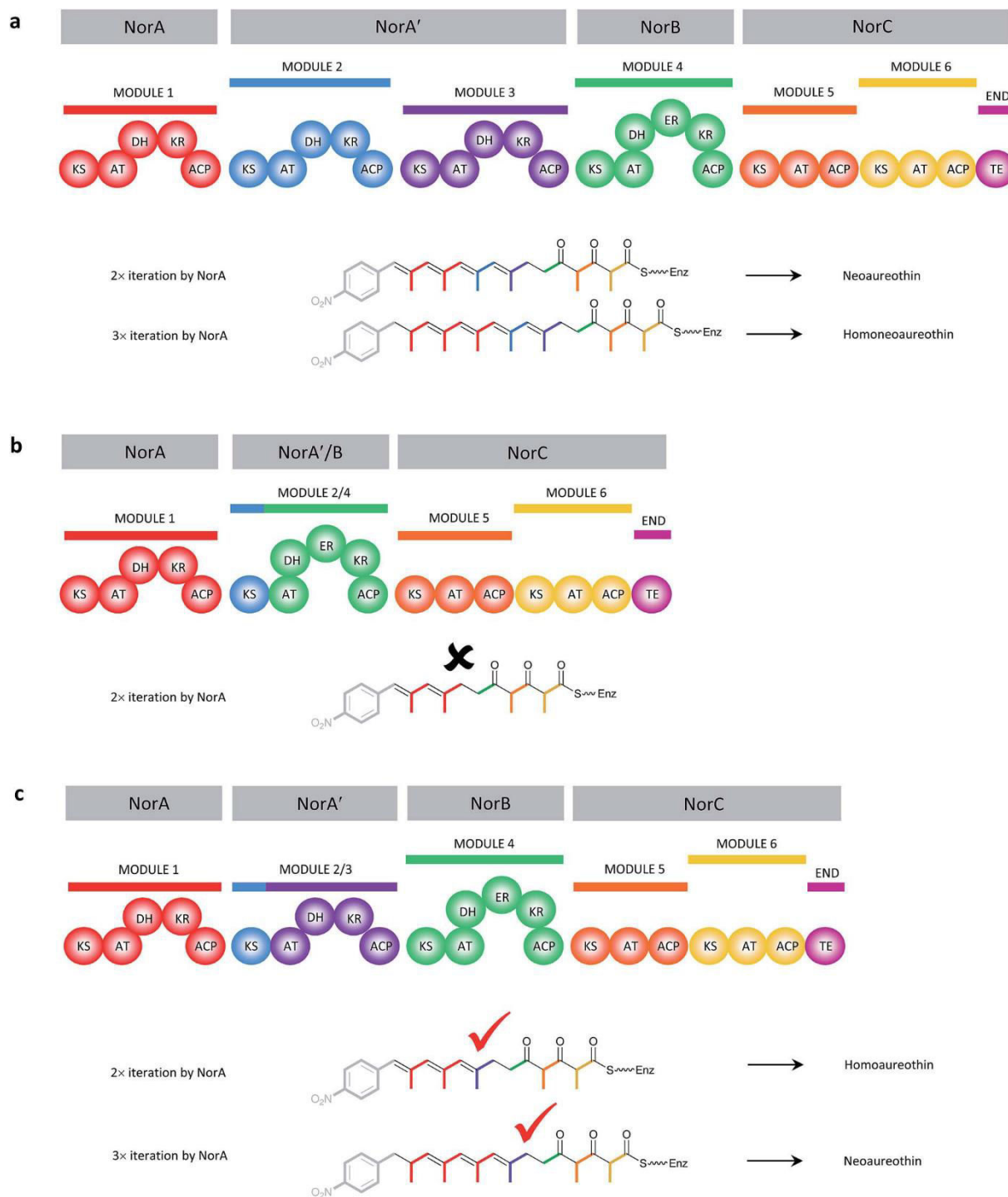


Figure 19 : Echange de module chez la PKS synthétisant néoauréothine (tiré de Weissman, 2016).

A : Schéma de la PKS native synthétisant la néoauréothine avec 2 itérations de NorA ou l'homonéoauréothine si NorA catalyse trois cycles d'élongation.

B : La KS du module 2 est conservée et fusionnée directement avec le module 4. Cet échange de module donne naissance à une PKS non fonctionnelle.

C : L'interface ACP1-KS2 est à nouveau conservée, et le reste des domaines du second module sont supprimés. Cette construction donne les deux produits attendus, plus courts de trois carbones que les deux produits naturels.

2. Les PKS de type *trans*-AT : une alternative intéressante pour créer de la diversité

Jusqu'à ce jour, peu d'expériences d'ingénierie des PKS de type *trans*-AT ont été menées. En effet, que ce soit le fonctionnement non canonique de ces PKS, ne respectant pas la règle de colinéarité, ou bien du fait du peu d'informations disponibles, ces systèmes ont été délaissés dans l'approche de biologie synthétique au profit des PKS de type *cis*. De plus, du fait de leur origine évolutive différente, nul ne sait si les informations collectées jusqu'à présent sur les PKS de type *cis* sont directement applicables aux PKS *trans*-AT. Cependant, l'architecture unique des PKS de type *trans*-AT les rend particulièrement attrayantes pour les approches de biologie synthétique (Till, et Race, 2016).

Tout d'abord, la particularité principale de ces systèmes présente beaucoup d'intérêt. En effet, bien qu'à l'heure actuelle peu d'informations structurales et mécanistiques ne soient disponibles sur les AT agissant en *trans*, ces protéines apparaissent plus attractives que leur équivalent chez les *cis*-AT. Ces protéines AT sont idéales pour l'expression recombinante, leur purification et donc leur caractérisation. De plus, les études menées jusqu'à présent ont démontré que les AT agissant en *trans* présentent le même repliement et les mêmes déterminants de la spécificité de substrat que les domaines AT agissant en *cis* (Wong, *et al.*, 2011 ; Till, et Race, 2014). Des AT peu spécifiques ont également été découvertes, permettant en théorie aux PKS qui utilisent ces AT de produire plusieurs polykétides différents selon la biodisponibilité des unités d'extension acyl-CoA (Koryakina, *et al.*, 2013).

Par ailleurs, de nombreuses PKS *trans*-AT comportent des modules NRPS. A toutes les modifications décrites sur les PKS de type *cis*, s'ajoutent donc des modifications de ces modules NRPS. Cela peut se faire, par exemple, au niveau de la sélection des acides aminés à incorporer (des acides aminés non-naturels peuvent par exemple être utilisés, bien que certains domaines A en incorporent également)). Des expériences de substitution ou de mutagenèse dirigée ont été menées sur le domaine d'adénylation et ont permis la génération de nouvelles molécules (Thirlway, *et al.*, 2012 ; Evans, *et al.*, 2011). Les domaines de modifications optionnelles des NRPS, présentés dans le point III.4.B)b. peuvent également être une cible potentielle d'ingénierie.

3. Une approche d'avenir, mais de nombreux points restent à éclaircir

La grande majorité des expériences d'ingénierie des PKS ont eu un succès mitigé (Weissman, 2016). Certaines n'ont pas permis d'obtenir le produit escompté, la synthèse aboutissant à la formation d'un produit abortif (Bevitt, *et al.*, 1993). D'autres tentatives ont débouché sur la synthèse de plusieurs molécules par une PKS, le polykétide escompté étant sous forme minoritaire (Rowe, *et al.*, 2001). Et pour finir, les approches ayant permis d'obtenir les polykétides désirés ont conduit à de fortes pertes de rendement. Les difficultés rencontrées proviennent vraisemblablement de la spécificité des domaines des modules en aval de la modification de la PKS. L'intégrité structurale de la PKS doit également être respectée et les interactions protéines/protéines au sein de la PKS doivent toujours être possibles.

Le succès de l'ingénierie des PKS réside dans l'accumulation de connaissances dans cinq grands domaines : i) étudier la spécificité de substrat des domaines et la façon dont ils interagissent pour contrôler la stéréochimie ; ii) comprendre le mécanisme par lequel la chaîne en élongation entre dans le site actif d'un domaine à l'autre, ce qui revient à étudier les interactions entre l'ACP et tous les autres domaines catalytiques présents au sein du même module ; iii) acquérir davantage de connaissances sur la structure et la dynamique des modules, et ce à plusieurs stades de la catalyse, afin d'appréhender les mécanismes nécessaires à la communication entre domaines au sein d'un module ; iv) caractériser la communication entre deux modules consécutifs, en particulier lorsque ces modules sont portés par deux chaînes peptidiques différentes, une interaction spécifique médiée par des domaines de docking (Broadhurst, *et al.*, 2003) ; et v) réaliser des études structure/fonction des enzymes post-PKS dans le but d'améliorer notre capacité à utiliser ces enzymes pour générer de nouveaux polykétides.

V. Etat des connaissances des structures et mécanismes des domaines PKS

Afin d'améliorer la connaissance des systèmes PKS, de nombreuses études structurales et mécanistiques ont été entreprises. Dans cette partie sera présenté les connaissances primordiales des relations structure/fonction sur chacun des domaines PKS.

1. Les domaines « acyl carrier protein »

Les domaines ACP constituent avec les *linkers* les seuls domaines non catalytiques des PKS. Il s'agit d'un domaine globulaire d'environ 80 résidus. Les ACP sont activés par une modification post-traductionnelle consistant en l'ajout d'un groupement phosphopantéthéine sur la sérine canonique du motif conservé "DSL". Ce groupement provenant du CoA est transféré par une enzyme appelée phosphopantéthéinyl transférase (PPTase), dont le gène est souvent présent au sein du cluster de gènes codant la PKS (Figure 20).

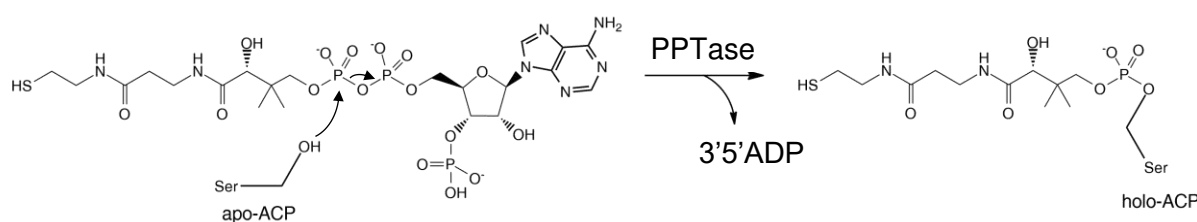


Figure 20 : Réaction de transfert du groupement phosphopantéthéine du coenzyme A sur la sérine de l'ACP catalysée par la PPTase (Lambalot, et al., 1996).

Le groupement phosphopantéthéine sert de support à la chaîne en élongation. Ce bras flexible, long d'environ 20 Å, permet l'entrée du polykétide dans les sites actifs des différents domaines.

Les ACP se structurent en fagot de quatre hélices α formées de trois hélices majeures et d'une courte hélice perpendiculaire aux trois précédentes (Figure 21). Leur structure caractéristique est très conservée entre les ACP/PCP. L'hélice I est la plus longue, et une longue boucle la sépare de l'hélice II. Cette dernière comporte la sérine canonique à son extrémité C-terminale. La courte hélice III est orthogonale aux hélices I, II et IV. Plusieurs études suggèrent que l'hélice II portant la sérine modifiée serait aussi impliquée dans les interactions protéine-protéine que l'ACP doit établir

avec les différents domaines (Weissman, *et al.*, 2006 ; Worsham, *et al.*, 2003), bien que quelquefois aucune interaction directe n'est requise du fait du lien covalent entre domaines (Tran, *et al.*, 2008).

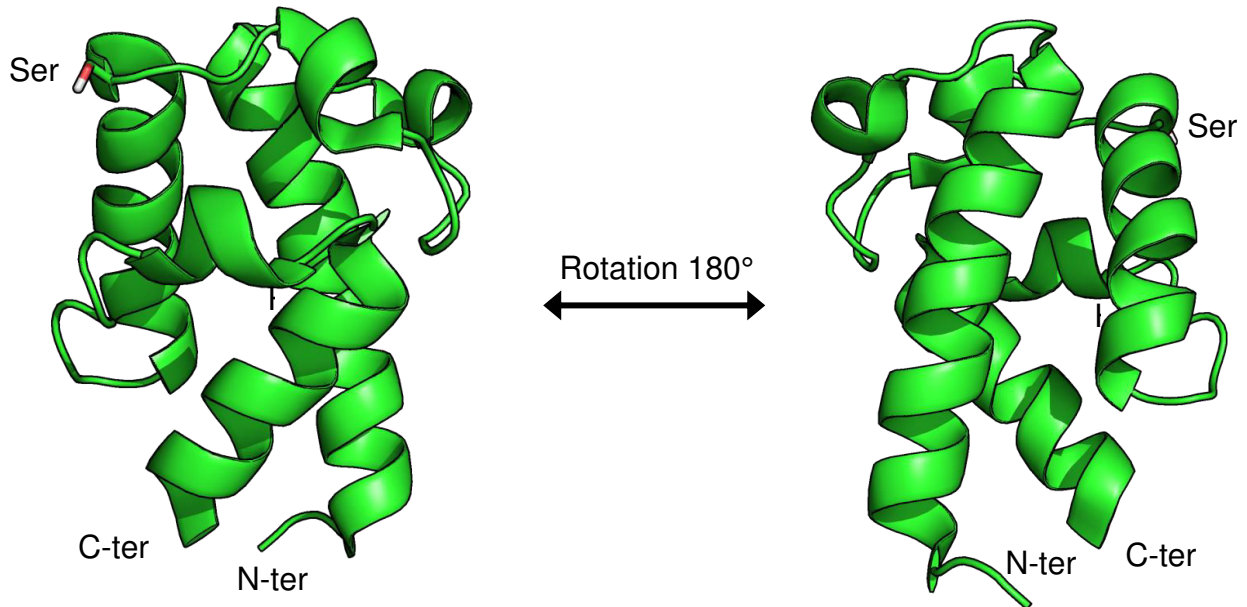


Figure 21 : Structure RMN de l'ACP du module 2 de DEBS (pdb : 2JU2 ; Alekseyev, *et al.*, 2007).

Les éléments de structure secondaire ainsi que la sérine canonique du motif DSL sont indiqués sur la structure.

2. Les domaines acyltransférase

Le domaine AT est responsable de la sélection des unités d'extension et de leur transfert sur le bras phosphopantéthéine de l'ACP. Ce domaine est présent au sein des modules chez les PKS *cis*-AT mais constitue une protéine discrète chez les types *trans*.

Ce sont des domaines monomériques, d'environ 300 acides aminés, reliés au KS par un linker « KS-AT » structuré d'environ 140 résidus. Les AT comprennent deux sous-domaines : un sous-domaine catalytique d'environ 240 acides aminés qui présente un repliement de type α/β -hydrolase, et un sous-domaine de 60 acides aminés qui a un repliement de type ferrédoxine (Figure 22). Les structures des didomaines KS-AT des modules 3 et 5 de DEBS ont été résolues par cristallographie (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007), de même que d'autres AT des systèmes FAS (Serre, *et al.*, 1995 ; Zhang, *et al.*, 2007 ; Pappenberger, *et al.*, 2010). Une structure

d'AT discrète, provenant de système *trans*-AT a également été résolue et présente le même repliement que les AT des PKS de type *cis* (Wong, *et al.*, 2011).

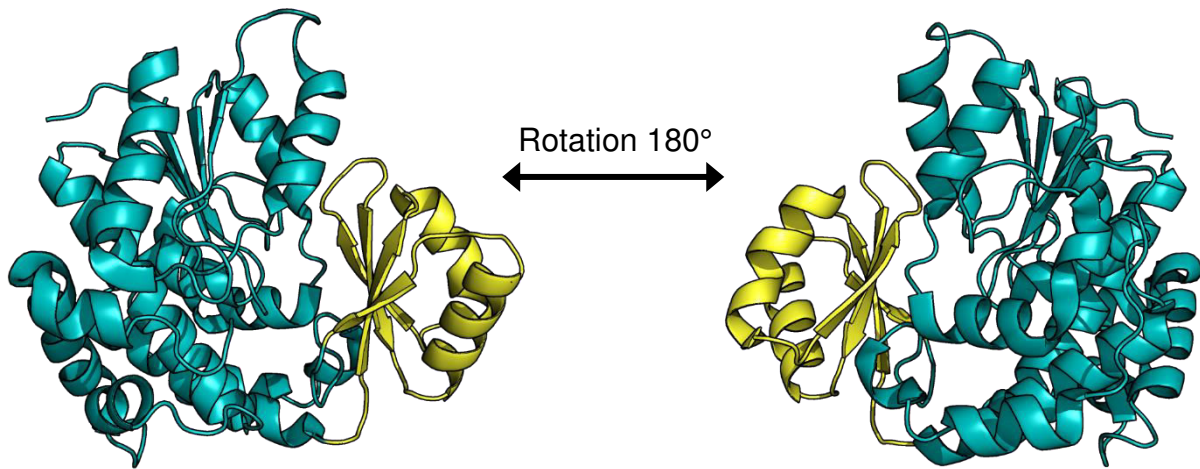


Figure 22 : Structure de l'AT du module 5 de DEBS (PDB : 2HG4 (Tang, *et al.*, 2006)). Ici sont détaillés les deux sous-domaines de l'AT : le sous-domaine catalytique avec un repliement de type α/β hydrolase représenté en bleu et le sous-domaine présentant un repliement de type ferrédoxine est représenté en jaune.

Le mécanisme catalytique de transfert de l'unité d'extension est semblable à celui d'une protéase à sérine. La sérine catalytique de l'AT est activée par une histidine à son tour activée par un carbonyle de la chaîne principale. La sérine réalise ainsi une attaque nucléophile sur l'acyl-CoA permettant ainsi le départ du CoA, puis le groupement thiol du bras phosphopantéthéine de l'ACP attaque l'intermédiaire AT-acyl (Figure 23).

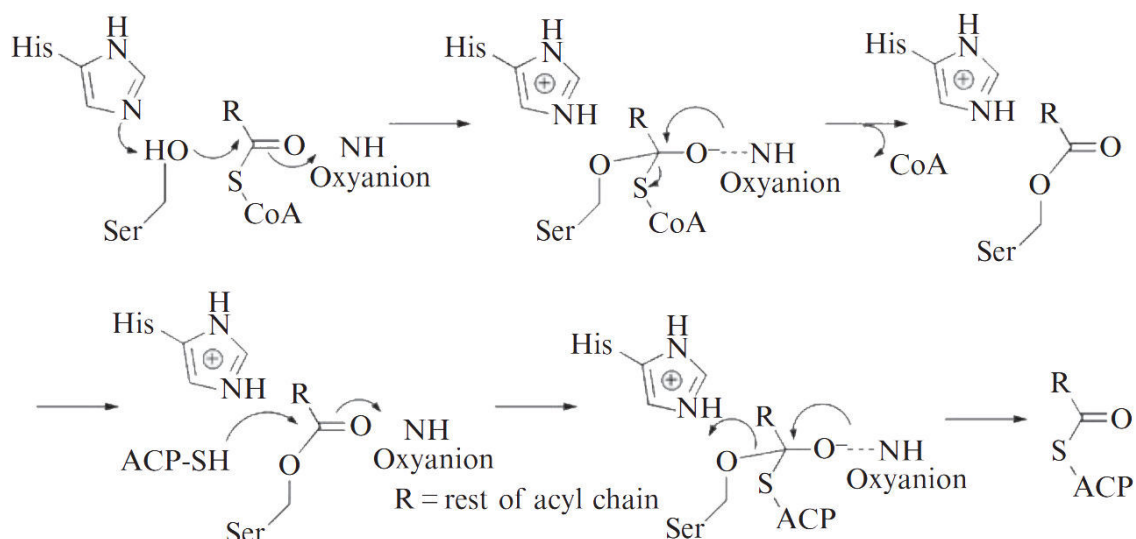


Figure 23 : Mécanisme de transfert d'unité d'extension sur l'ACP catalysé par l'AT (Tsai, et Ames, 2009).

L'histidine catalytique active la sérine catalytique par une catalyse base. La sérine réalise alors une attaque nucléophile au niveau du groupement thioester, provoquant ainsi le départ du CoA. Ensuite, le groupement thiol du bras phosphopantéthéine de l'ACP attaque l'unité d'extension au niveau du même carbone, réaction catalysée par une catalyse acide de l'histidine permettant la régénération de la sérine catalytique. Il est à noter que les charges sur les intermédiaires sont stabilisées par un des NH peptidiques.

La spécificité des AT pour les unités d'extension est corrélée à des motifs présents au sein du site actif (Haydock, *et al.*, 1995). Les AT spécifiques pour le malonyl-CoA présentent des motifs **GHS(I/V)G** et **HAFH** alors que les AT spécifiques pour le (2S)-méthylmalonyl-CoA (Marsden, *et al.*, 1994) arborent un motif **GHSQG** et **YASH** où les résidus catalytiques sont indiqués en gras (Yadav, *et al.*, 2003 ; Starcevic, *et al.*, 2008). Cependant, des mutations de ces motifs dans le but d'altérer la spécificité n'ont pas entraîné un changement complet dans la sélection de l'unité d'extension, indiquant que d'autres éléments contribuent à la spécificité des AT (Del Vecchio, *et al.*, 2003). Les motifs permettant la sélection des autres unités d'extension restent encore à être déterminés, ce qui pourrait permettre en théorie la conversion de la spécificité des AT par des approches rationnelles. Ceci s'avère toutefois relativement compliqué puisque la spécificité semble liée à plusieurs motifs.

3. Les domaines kétosynthase

Les domaines KS catalysent la formation de liaison carbone-carbone entre l'unité d'extension et la chaîne en élongation. Cette réaction est catalysée par des domaines d'environ 430 résidus. Les KS ont un repliement typique des thiolases avec une architecture $\alpha/\beta/\alpha/\beta/\alpha$. Ils sont dimériques, avec une interface d'interaction étendue entre les deux monomères qui est principalement assurée par l'intermédiaire de deux feuillets β (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007 ; Bretschneider, *et al.*, 2013) (Figure 24).

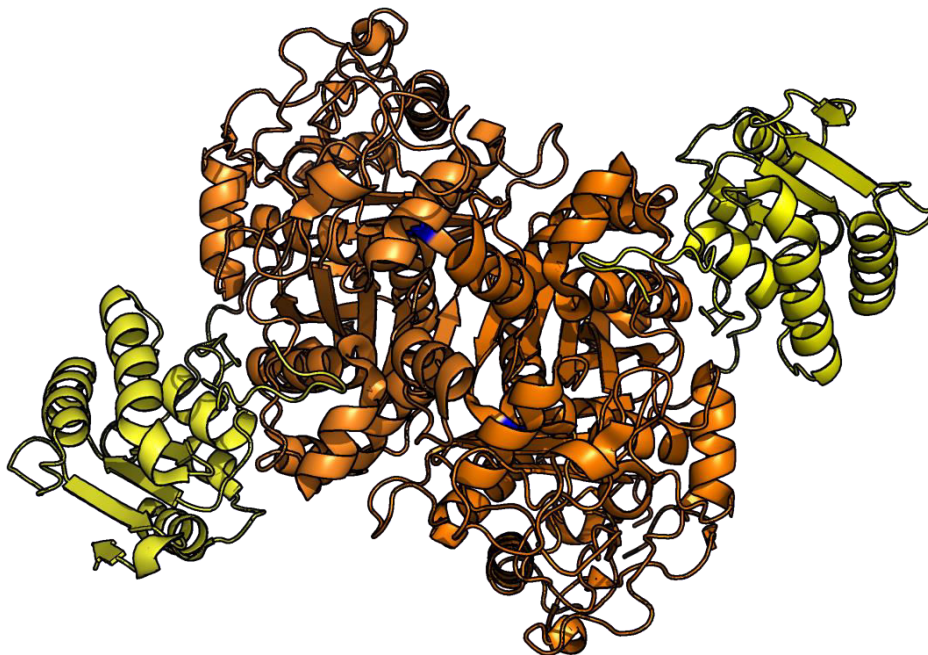


Figure 24 : Structure du KS-linker (PDB : 4KC5, Bretschneider, *et al.*, 2013).

Le dimère de KS est représenté en orange et la cystéine catalytique en bleu. Bien qu'il s'agisse d'une KS de PKS de type *trans*-AT, le linker séparant habituellement le KS de l'AT est tout de même présent et est représenté en jaune. Il est proposé que ce linker puisse servir d'ancrage à l'AT devant interagir en *trans* avec le domaine ACP (Tang, *et al.*, 2004).

Le mécanisme catalytique a été déterminé pour la KS de la FAS animal. Néanmoins, étant donnée la forte homologie avec les domaines KS des PKS, il est fort probable pour qu'il soit identique chez les PKS. La réaction est permise grâce à une triade catalytique composée d'une cystéine d'un motif TxCSS et de deux histidines des motifs HGTGT et KSNIGHT. La cystéine se situe à l'extrémité d'une longue hélice α , une autre hélice pointant également sur elle, ce qui permet de diminuer le pK_a et donc

d'activer la cystéine. Celle-ci peut donc attaquer la chaîne en élongation alors située sur le groupement phosphopantéthéine de l'ACP du module précédent. Une fois acylée, la KS catalyse la décarboxylation de l'unité d'extension portée par l'ACP du même module que le KS (Witkowski, *et al.*, 2002). Il est à noter que la mutation de la cystéine catalytique en glutamine mime l'état acylé du domaine KS et par conséquent, active son activité décarboxylase (Witkowski, *et al.*, 1999).

La décarboxylation de l'unité d'extension permet la création d'un carbone nucléophile qui réagit alors avec le carbone électrophile de la fonction thioester de la chaîne en élongation liée à la cystéine du KS, permettant la régénération de cette dernière (Figure 25).

Des études par microscopie électronique à transmissions de *Bacillus subtilis* ont révélé des masses denses à proximité de la membrane plasmique. Ces masses semblables à des organelles correspondraient à des amas de PKS synthétisant la bacillaene, un inhibiteur de la synthèse protéique chez les eucaryotes (Straight, *et al.*, 2007). Récemment, la résolution de 9 structures de domaines KS provenant de divers PKS *trans*-AT a mis en évidence un motif appelé LINKS (Laterally-Interacting Ketosynthase Sequence) (Gay, *et al.*, 2016). Ce motif serait situé au sein du linker KS-AT, qui sert d'adaptateur pour permettre l'interaction de la protéine AT avec les différents modules à acyler. Le motif LINKS au sein de ce domaine linker permettrait l'interaction entre les différents KS *via* des interactions ioniques et hydrophobes, ce qui permettrait la formation des mégacomplexes décrit par microscopie chez *B. subtilis*. Cependant, ces interactions n'ont été identifiées que dans les structures cristallographiques et pourraient représenter un artefact de cristallisation.

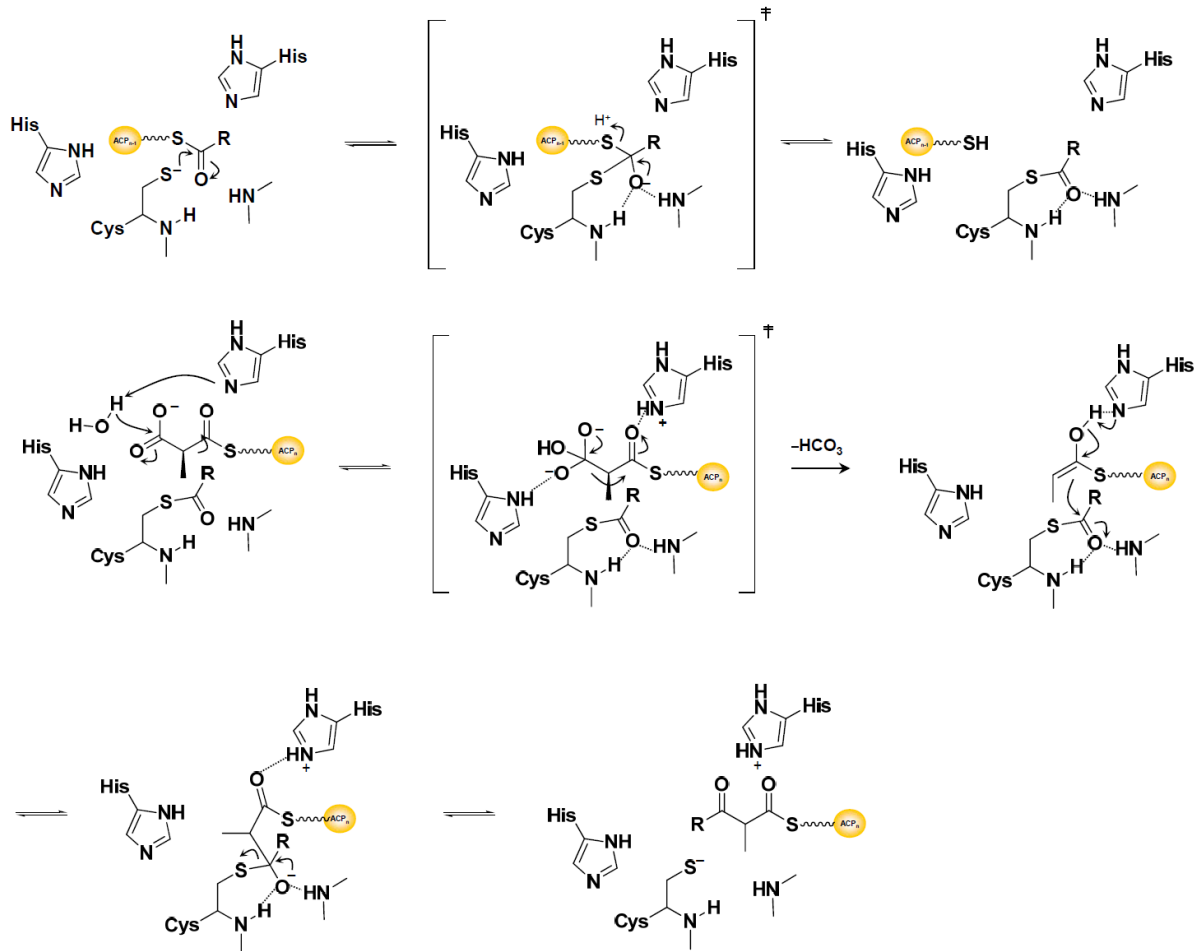


Figure 25 : Réaction de condensation de type Claisen catalysée par le domaine KS (d'après Witkowski, et al., 2002, adapté par KJ Weissman).

La cystéine activée attaque le carbone électrophile de l'unité d'extension porté par l'ACP_{n-1}. Une fois acylée, une histidine va jouer le rôle de catalyseur basique générale pour activer une molécule d'eau, qui va dès lors attaquer le groupement carboxylate de l'unité d'extension porté par l'ACP_n. L'intermédiaire tétraédrique résultant est alors stabilisé par les deux histidines, et s'ensuit une décarboxylation avec libération de bicarbonate. Le carbone nucléophile formé par la décarboxylation va attaquer le groupement thioester de la chaîne en élongation porté par la cystéine du KS afin de former une liaison carbone-carbone. La chaîne en élongation, ainsi allongée d'une unité acyle, se retrouve alors sur le bras phosphopantéthéine de l'ACP_n.

4. Les domaines kétoréductase

Les domaines KR catalysent la réaction d'oxydation de la fonction cétone, située en position β du polykétide en cours de synthèse, en hydroxyle. Ces domaines sont responsables de la stéréochimie des polykétide puisque la réduction du groupement β -céto en β -hydroxy est stéréospécifique. On note en effet deux types de KR, le type A qui conduit à la formation d'un hydroxyle de configuration *S* et le type B qui catalyse la formation d'un hydroxyle de configuration *R* (Caffrey, 2003). En outre, certaines KR catalysent également l'épimérisation des substituants en α (Keatinge-Clay, 2007 ; Garg, *et al.*, 2014). La direction d'entrée du substrat dans le site actif de la KR et sa position par rapport au NADPH serait à la base de la stéréochimie du carbone β (Siskos, *et al.*, 2005). Le contrôle de la stéréochimie du carbone α a soulevé davantage de questions. En effet, trois domaines interviennent dans la stéréochimie des substituants en α : le domaine AT de par la sélection d'unités d'extension stéréospécifiques (le (2*S*)-méthylmalonyl-CoA est toujours sélectionné et non l'isomère (2*R*) (Marsden, *et al.*, 1994)), le domaine KS qui catalyse la condensation de Claisen avec une inversion de configuration (Weissman, *et al.*, 1997), et enfin le domaine KR. Effectivement, certaines KR catalyse une réaction d'épimérisation, modifiant la configuration D-méthyl résultante de la condensation en L-méthyl (Garg, *et al.*, 2013 ; Garg, *et al.*, 2014). Cependant, aucune preuve au niveau moléculaire n'a encore été établie quant au mécanisme de cette épimérisation, même si la tyrosine et la sérine catalysant la réduction semblent également impliquée dans l'épimérisation (Xie, *et al.*, 2016). Les KR sont donc rangées en 6 classes, selon leur stéréospécificité, leur activité épimérase, voire selon leur absence d'activité réductase : les classes A1 et B1 qui ne présentent pas d'activité épimérase, A2 et B2 qui ont une activité épimérase, C2 qui a simplement une activité épimérase et C1 qui est complètement inactives à cause d'une mutation dans le motif de fixation du NADPH et enfin, A0 et B0 lorsqu'aucun substituant n'est présent en position α (Keatinge-Clay, 2007).

Les KR sont des domaines monomériques, d'environ 450 acides aminés. Ils se composent de deux sous-domaines : un domaine catalytique d'environ 250 acides aminés et un domaine structural d'environ 200 résidus. Ces deux sous-domaines sont relativement identiques, étant tout deux composés d'un Rossmann *fold* et présentant un repliement de type déshydrogénase/réductase à courte chaîne (SDR). Par

conséquent les KR sont souvent considérés comme étant des pseudo-dimères (Figure 26).

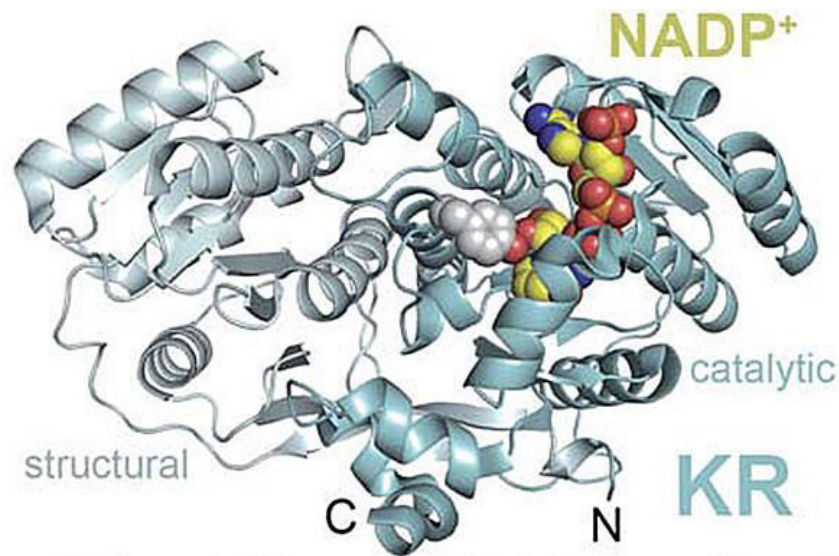


Figure 26 : Structure du domaine KR du module 2 de la PKS synthétisant l'amphotéricine (PDB : 3MJS (Zheng, et al., 2010), tirée de Keatinge-clay, 2012).

Le domaine KR est composé d'un domaine structural (en gris) et d'un domaine catalytique capable de fixer du NADPH/NADP⁺.

La catalyse implique une tétrade catalytique composée d'une tyrosine qui joue le rôle de catalyseur acide général. Une lysine et une glutamine servent de relais de protons impliquant également le co-facteur NADPH et une molécule d'eau, ce qui permet d'abaisser le pK_a de la tyrosine. Enfin, une sérine permet de stabiliser la charge négative qui se développe sur l'oxygène du polykétide durant la catalyse, et d'augmenter le caractère électrophile du groupement cétone (Tsai, et Ames, 2009 ; Keatinge-Clay, et Stroud, 2006 ; Keatinge-Clay, 2007) (Figure 27).

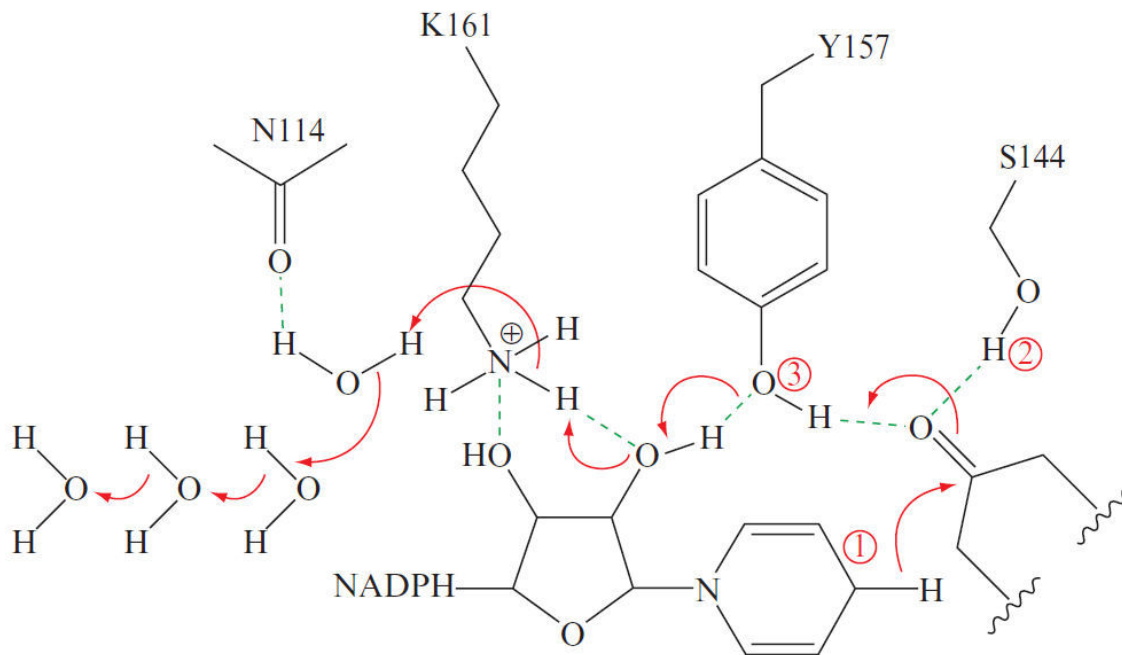


Figure 27 : Mécanisme catalytique des domaines KR (tiré de Tsai, et Ames, 2009).

Le NADPH transfère un hydrure au niveau du carbone β du polykétide en élévation porté par l'ACP. Une tyrosine du domaine KR joue alors le rôle de catalyseur acide et cède un proton pour stabiliser l'oxyanion généré par la réduction. Cette tyrosine sera ensuite régénérée par un relai de proton impliquant une lysine, une glutamine, l'hydroxyle du ribose, ainsi que quatre molécules d'eau.

5. Les domaines déshydratase

Le domaine DH intervient suite à la réduction de la cétone en hydroxyle par la KR. Il catalyse la déshydratation de l'hydroxyle du polykétide en élévation afin de former une double liaison entre les carbones α et β . Ce sont des domaines dimériques d'environ 280 résidus qui présente un repliement dit en « double hot dog » dans lequel deux hélices centrales sont flanquées de deux feuillets β (Dillon et Bateman, 2004 ; Keatinge-Clay, 2008 ; Akey, *et al.*, 2010 ; Gay, *et al.*, 2013) (Figure 28).

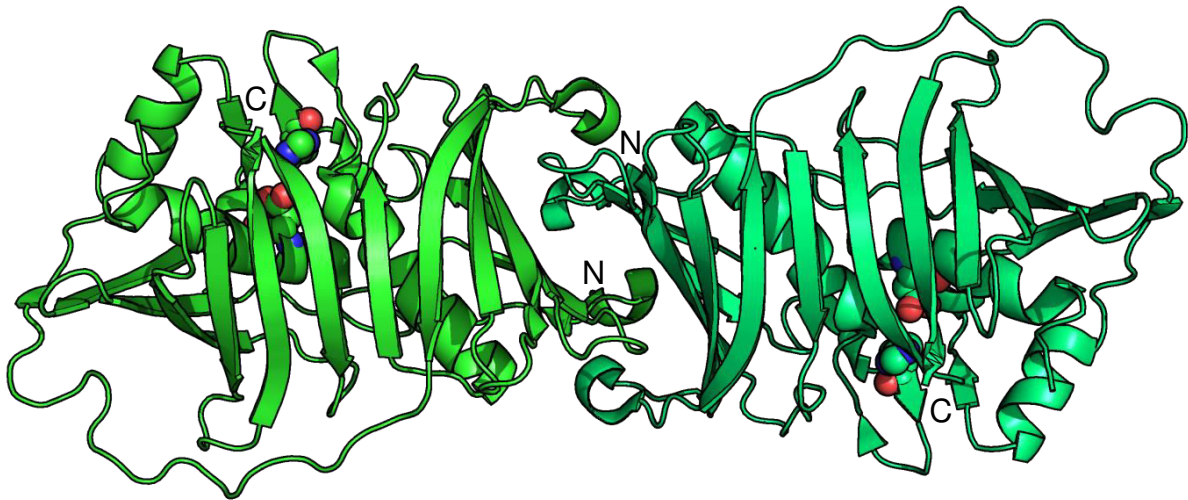


Figure 28 : Structure du domaine DH du module 4 de DEBS (PDB : 3EL6 (Keatinge-Clay, 2008)).

Les domaines DH présentent un repliement en double hot-dog dans lequel les deux hélices centrales forment la saucisse et les deux feuillets flanquant, le pain. Les résidus aspartate et histidine impliqués dans la catalyse sont représentés en sphère.

Au vu des structures, deux résidus seraient impliqués dans la catalyse : une histidine, qui va jouer le rôle de catalyseur base pour capter un hydrogène du carbone α , et un aspartate, qui agit en tant que catalyseur acide général, pour transmettre un proton au groupement hydroxyle, afin de faciliter son départ sous forme d'une molécule d'eau (Leesong, *et al.*, 1996 ; Keatinge-Clay, 2008) (Figure 29). Cependant, sur la base de mutagenèses dirigées, il semblerait que seule l'histidine serait nécessaire à la catalyse (Labonte, et Townsend, 2013). La structure d'une hydratase, qui catalyse donc la réaction inverse, mais avec le même repliement en double hot-dog que les DH, semble également indiquer que seule l'histidine jouerait le rôle de catalyseur à la fois base et acide (Koski, *et al.*, 2004).

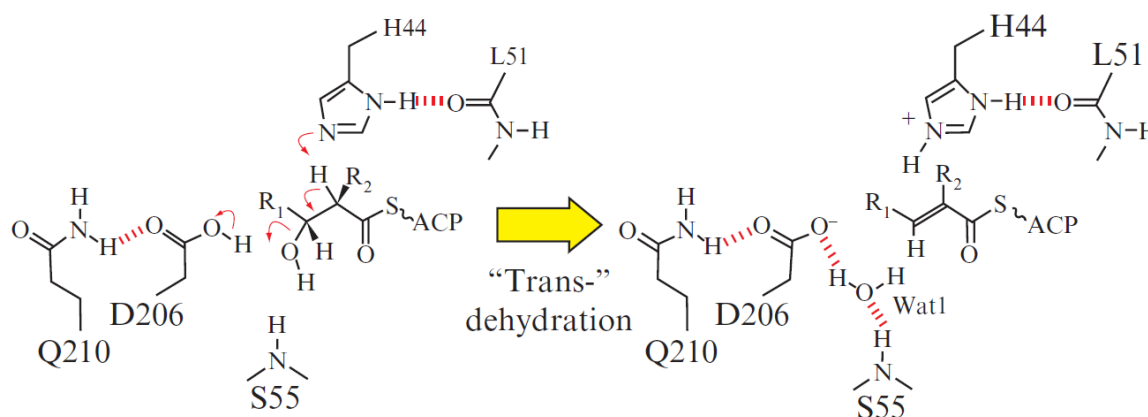


Figure 29 : Mécanisme catalytique proposé pour le domaine DH du module 4 de DEBS (tiré de Tsai et Ames, 2009).

L'His44 joue le rôle de catalyseur basique général pour déprotoner le carbone α du polykétide lié à l'ACP. L'Asp206 sert de catalyseur acide général pour transmettre un proton à l'hydroxyle pour favoriser son élimination sous forme d' H_2O . Le dipôle créé par l'hélice 1 polarise également le groupement hydroxyle, ce qui facilite son élimination. La stéréochimie du carbone β serait en partie contrôlée par le domaine KR, ce qui induit la configuration *cis* ou *trans* de l'énoyl résultant ou pourrait être dû à un mécanisme d'entrée du substrat différent dans les deux cas. Une isomérisation d'une double liaison *trans* pourrait également expliquer l'apparition de doubles liaisons *cis*. (KJ. Weissman, à paraître)

6. Les domaines énoyl réductase

Les domaines ER interviennent dans la réduction stéréospécifique de la double liaison *trans*- α,β formée par le domaine DH. Ce sont les domaines les moins étudiés à l'heure actuelle, en dehors des domaines MT, vraisemblablement à cause de leur faible fréquence d'apparition au sein des modules PKS. Ce sont des domaines d'environ 310 acides aminés qui appartiennent à la famille des déshydrogénase/réductase à courte chaîne (MDR), et qui sont des enzymes à cofacteur NADP(H).

Les domaines ER présentent la particularité d'être dimérique chez les FAS et seraient monomériques chez les PKS (Zheng, *et al.*, 2012). Ils sont composés de deux sous-domaines : un sous domaine de fixation de nucléotide, d'environ 140 acides aminés se structurant en Rossmann *fold*, qui va permettre de fixer le co-facteur NADPH, et un sous-domaine de fixation du substrat d'environ 170 résidus (Zheng, *et al.*, 2012 ; Ames, *et al.*, 2012) (Figure 30).

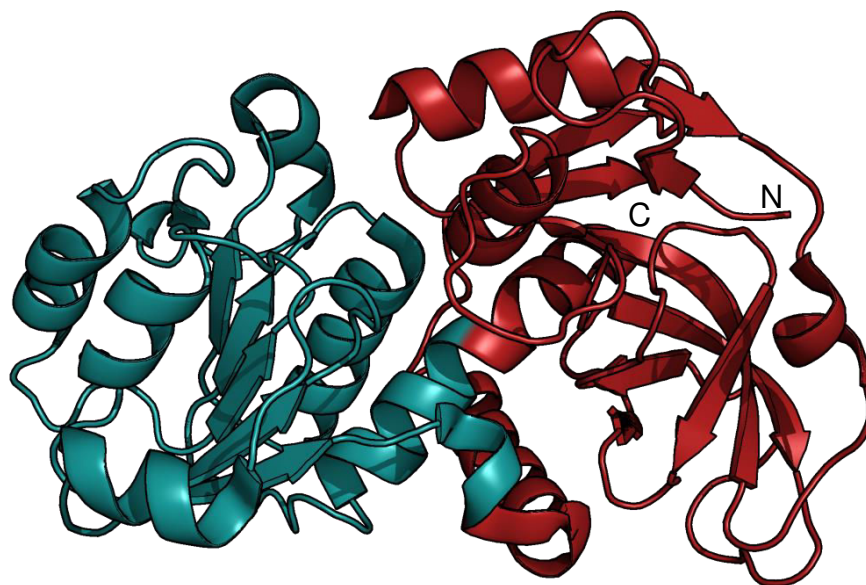


Figure 30 : Structure du domaine ER de la PKS synthétisant la spinosyne (PDB : 3SLK : Zheng, et al., 2012).

Le sous-domaine de fixation du NADP(H) est représenté en bleu et le sous-domaine de fixation du substrat en rouge. Le site actif de l'ER se situe à l'interface entre les deux sous-domaines.

La catalyse implique un résidu aspartate, et une lysine ou une tyrosine. L'implication de l'un ou l'autre de ces résidus va contrôler la stéréochimie de la réaction donnant naissance à un carbone α avec un substituant en orientation D ou L (Kwan, et al., 2008 ; Kwan, et Leadlay, 2010 ; Zheng, et al., 2010) (Figure 31).

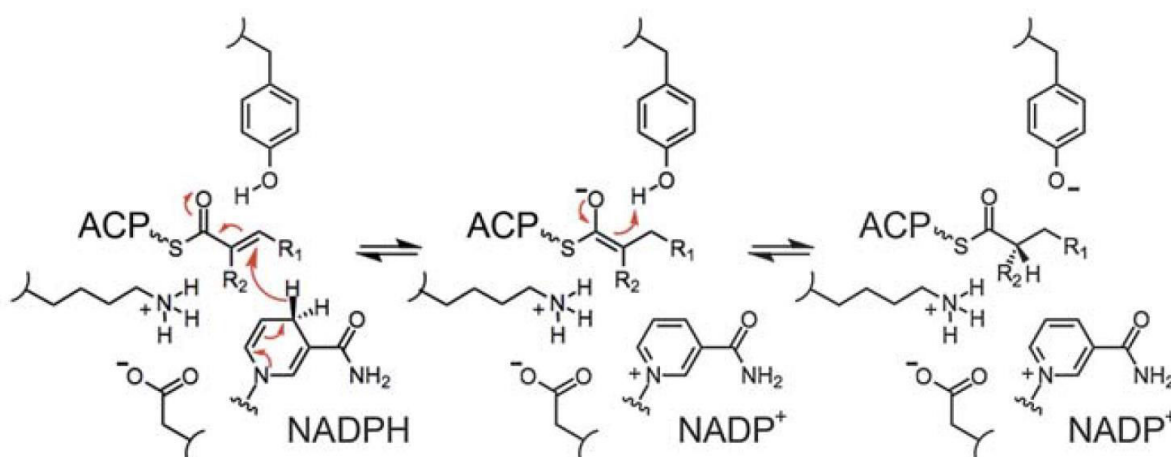


Figure 31 : Mécanisme catalytique des domaines ER (tiré de Keatinge-Clay, 2012).

Le NADPH va transférer un hydruure au niveau du carbone β du polykétide porté par l'ACP afin de former un intermédiaire énolate. Ensuite la tyrosine (ou la lysine située de l'autre côté du substrat) va jouer le rôle de catalyseur acide pour céder un proton au carbone α ce qui va permettre de former un acyl saturé.

7. Les domaines méthyltransférase

Ces domaines monomériques d'environ 320 acides aminés catalysent la réaction de transfert d'un groupement méthyle du co-facteur SAM vers le carbone α du polykétide. Ces domaines sont extrêmement rares chez les PKS de type *cis*-AT. En revanche, ils permettent de compenser le manque de diversité dans le choix des unités d'extension chez les PKS itératives et les PKS de type *trans*-AT, chez lesquels il est fréquent de retrouver ce domaine.

Il est composé de deux sous-domaines : un sous-domaine de 130 acides aminés en N-terminal qui lie le substrat et un domaine de 190 résidus en C-terminal qui fixe le co-facteur SAM (Vidgren, *et al.*, 1994 ; Maier, *et al.*, 2008) (Figure 32). A ce jour, seul des données structurales sur les domaines MT de FAS sont disponibles, aucune structure de ce domaine chez les PKS n'ayant encore été résolue.

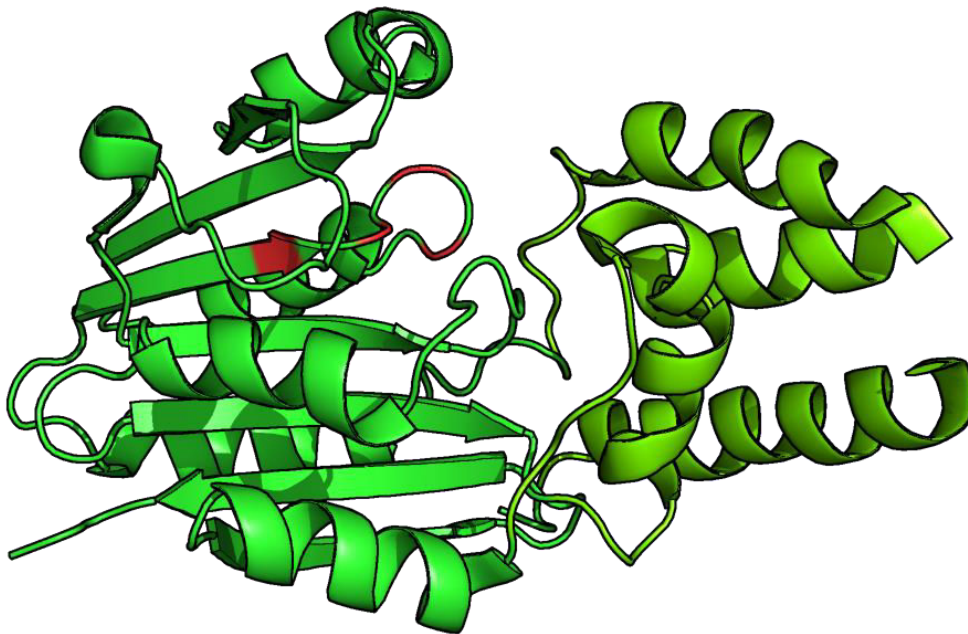


Figure 32 : Structure du domaine MT de la FAS porcine (PDB : 2VZ9 ; Maier, *et al.*, 2008). A gauche, coloré en vert, se trouve le domaine de fixation du SAM (le motif conservé D/ExGxGxG est coloré en rouge), et à droite coloré en vert pistache, se trouve le domaine de fixation du substrat.

Le mécanisme de transfert exact du groupement méthyle depuis le SAM jusqu'au polykétide en extension n'est pas encore connu, de même que le mécanisme d'activation du carbone α du polykétide qui doit subir une déprotonation pour le rendre

nucléophile (Figure 33). De même, la méthylation pourrait se faire sur l'unité d'extension attachée à l'ACP ou bien consécutivement à la réaction de condensation catalysée par le domaine KS. Cette dernière voie semble être la bonne, puisque, en plus de présenter une chimie plus favorable (le proton sur le carbone α est plus acide suite à la condensation), des expériences de LC-MS n'ont pas permis de mettre en évidence une augmentation de masse du malonyl-ACP mais ont permis au contraire d'identifier une méthylation du β -kétacyl-ACP (Wagner, *et al.*, 2016).

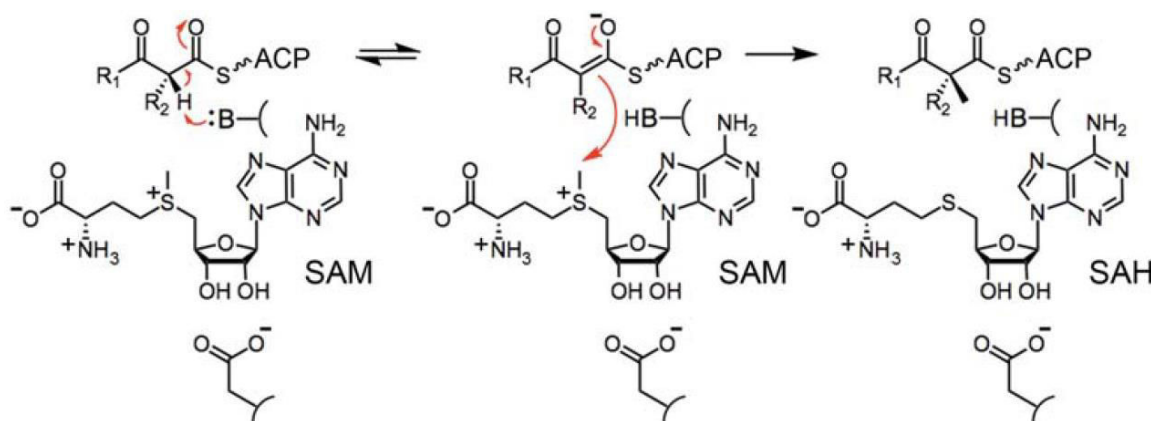


Figure 33 : Mécanisme putatif d' α -méthylation catalysée par les domaines MT (tiré de Keatinge-Clay, 2012).

Un catalyseur basique général viendrait capter un proton au niveau du carbone α pour activer son caractère nucléophile, ce qui lui permettrait de réagir avec le centre électrophile du SAM et ainsi de récupérer le groupement méthyle.

8. Les domaines thioestérase

Ce sont des domaines de 240 à 290 acides aminés généralement fusionnés à l'extrémité C-terminale du dernier module de la PKS. La plupart des domaines TE catalysent la cyclisation intramoléculaire des polykétides afin de libérer la chaîne en fin de synthèse pour former une macrolactone. En revanche certains TE catalysent la libération du polykétide par hydrolyse, ce qui génère alors un polykétide linéaire qui peut être ensuite cyclisé par une enzyme post-PKS.

Certains type de TE (appelés TE de type II) constituent des protéines discrètes agissant en *trans* pour catalyser l'hydrolyse de produits aberrants qui serait susceptible de bloquer la machinerie de synthèse de la PKS. Ces TE, aussi appelés « proof-reading », permettent ainsi de régénérer la PKS qui serait sinon inhibée (Heathcote, *et al.*, 2001 ; Schwarzer, *et al.*, 2002 ; Yeh, *et al.*, 2004).

Les domaines TE des PKS de type I modulaires sont dimériques (Tsai, *et al.*, 2001 ; Tsai, *et al.*, 2002 ; Scaglione, *et al.*, 2010) alors que ceux des PKS de type I itératives et des FAS semblent monomériques (Chakravarty, *et al.*, 2004). Ils présentent un repliement de type α/β hydrolase composé de huit feuillets β et de six hélices α . Un long canal d'entrée du substrat traverse la protéine de part en part. Le site actif, situé au milieu de ce canal, comprend une triade catalytique composée d'un aspartate, une histidine et une sérine. Une longue boucle en périphérie forme un couvercle permettant la liaison du substrat. De plus, chez les PKS de type I modulaires, deux hélices en N-terminal riches en leucines permettent la dimérisation des TE (Figure 34) (Tsai, *et al.*, 2001 ; Tsai, *et al.*, 2002 ; Scaglione, *et al.*, 2010).

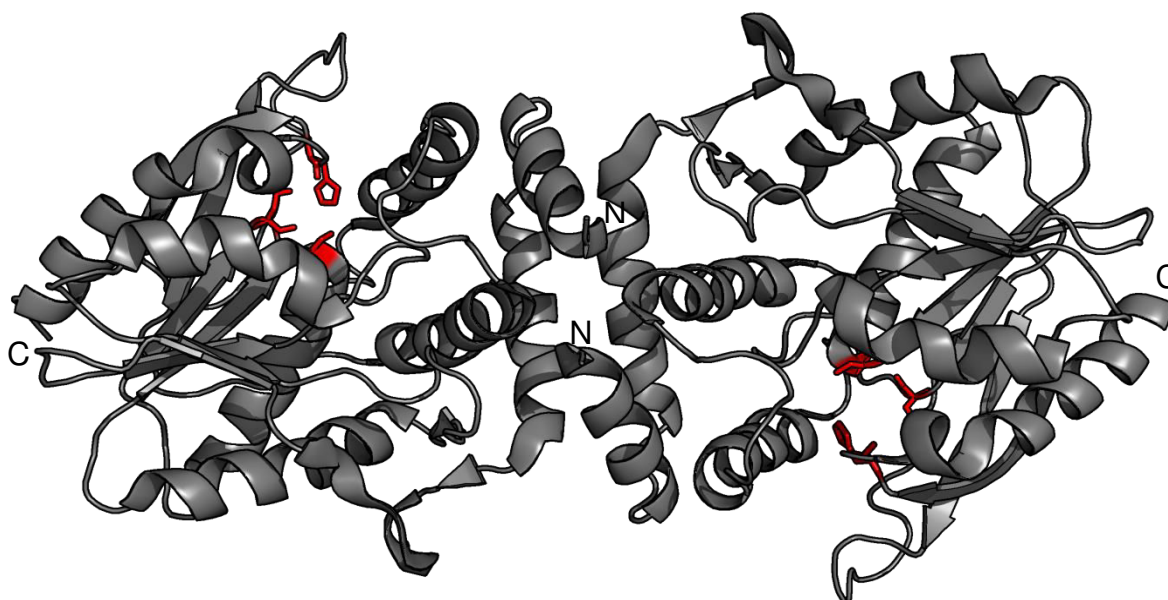


Figure 34 : Structure du domaine TE de DEBS (PDB : 1KEZ ; Tsai, *et al.*, 2001).

Le domaine TE se dimérise par l'intermédiaire de deux hélices α en N-terminale. Les résidus Ser, His et Asp de la triade catalytique sont indiqués en rouge. La boucle à proximité du site actif sert de site de liaison au substrat.

Le mécanisme de formation de la macrolactone est similaire à celui des domaines AT, à l'exception de la deuxième étape puisque l'attaque nucléophile sur l'intermédiaire acyl-enzyme ne se fera pas par le thiol du groupement phosphopantéthéine mais par un hydroxyle du polykétide (Figure 35).

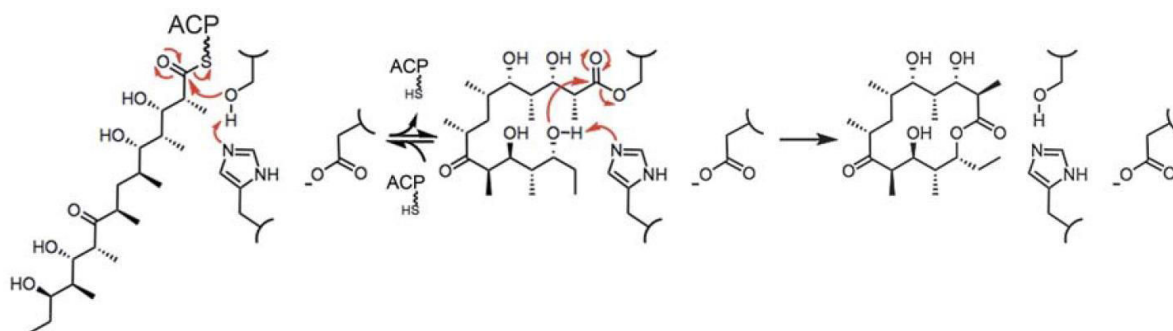


Figure 35 : Mécanisme catalytique général des domaines TE (tiré de Keatinge-Clay, 2012).

La sérine catalytique, activée par l'histidine, réalise une attaque nucléophile au niveau du groupement thioester formé par le polykétide lié à l'ACP. En règle générale, pour la seconde étape, l'histidine va cette fois activer un groupement hydroxyle préalablement formé au cours de la synthèse. Ce groupement va alors à son tour réaliser une attaque nucléophile au niveau de l'intermédiaire acyl-enzyme entraînant ainsi la macrocyclisation du polykétide.

VI. Organisation structurale des modules PKS

Il est clair qu'une structure tridimensionnelle d'un module ou d'une sous-unité multi-modulaire pourrait révolutionner notre compréhension des systèmes PKS. Pourtant, peu d'informations structurales sur les PKS sont disponibles à ce jour, et ceci est encore plus véridique pour le type *trans*-AT. En effet, aucune structure cristallographique d'un module PKS n'a pu être obtenue à ce jour, vraisemblablement à cause de la flexibilité intrinsèque de ces protéines. Cependant, une structure d'une PKS itérative a également été résolue récemment (Herbst, *et al.*, 2016).

Dès lors, la structure de la mFAS porcine résolue à 3.2 Å de résolution (voir II.2) a fourni de précieuses informations au niveau structural sur les PKS de type I.

1. La structure de la mFAS comme modèle pour les PKS de type I ?

En effet, du fait de leur forte homologie, on serait en droit de penser que la structure d'un module de PKS de type I – et plus particulièrement un module comportant le panel entier de domaines réducteurs – serait identique à celle de la mFAS.

Les structures des domaines de PKS résolues à ce jour vont dans ce sens puisque la plupart d'entre elles sont sensiblement identiques pour les deux synthèses. C'est également le cas des structures du didomaine KS-AT des modules 3 et 5 DEBS résolue à 2.7 Å de résolution (Figure 36) (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007). La structure des deux domaines, ainsi que la topologie du didomaine sont quasiment identiques à celles retrouvées dans la portion catalysant la condensation chez la mFAS (KS-MAT). Une superposition au niveau du KS des deux structures ne montre qu'un décalage de 10° de la position de l'AT par rapport au domaine KS.

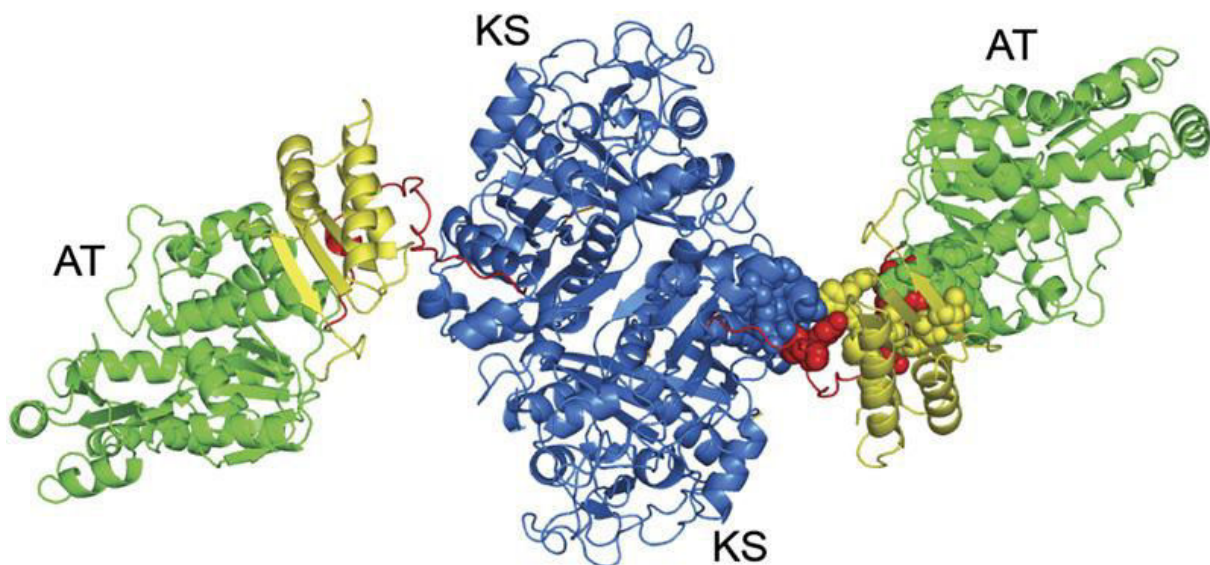


Figure 36 : Structure du KS-AT du module 3 de DEBS (tiré de Tang, *et al.*, 2007).

Le domaine KS dimérique occupe une place centrale (bleu). Le domaine AT se situe de part et d'autre du dimère de KS (vert). Un linker structuré (aussi appelé domaine adaptateur) relie le KS à l'AT (jaune) et est également constitué d'une portion du linker (en rouge) reliant l'AT au domaine suivant.

L'organisation du domaine KR est également similaire chez les deux synthèses. En effet, pour ce dernier, on retrouve un domaine KR catalytique et un domaine ψ KR qui présente une fonction plutôt plus structurale, et ceci à la fois chez les PKS (Keatinge-Clay, et Stroud, 2006) et chez la mFAS (Maier, *et al.*, 2008).

En revanche, la résolution de structures d'autres didomaines PKS a permis de mettre en évidence des différences entre les deux synthases. C'est notamment le cas pour le domaine ER qui est dimérique chez la mFAS et qui est clairement monomérique dans la structure du didomaine KR-ER (Zheng, *et al.*, 2012). *A contrario*, le domaine TE est monomérique chez les FAS et se trouve être dimérique chez les PKS de type I modulaires (Tsai, *et al.*, 2001 ; Chakravarty, *et al.*, 2004 ; Akey, *et al.*, 2006). De plus, l'interface de dimérisation des DH de PKS est également différente de celle des DH de FAS (Keatinge-Clay, 2008 ; Akey, *et al.*, 2010). Ces différences, en plus du fait que les PKS sont des systèmes multimodulaires et non itératifs comme la FAS, ont alors permis de proposer un nouveau modèle pour les PKS de type I (Zheng, *et al.*, 2012).

Dans ce nouveau modèle, l'interface de dimérisation plus « aplatie » du domaine DH ainsi que l'état oligomérique du domaine ER permettent de faciliter l'accès de l'ACP à ces domaines, d'autant plus que l'ACP, par rapport aux FAS, doit à la fois pouvoir interagir avec les domaines du même module mais également avec le KS du module suivant (Figure 37).

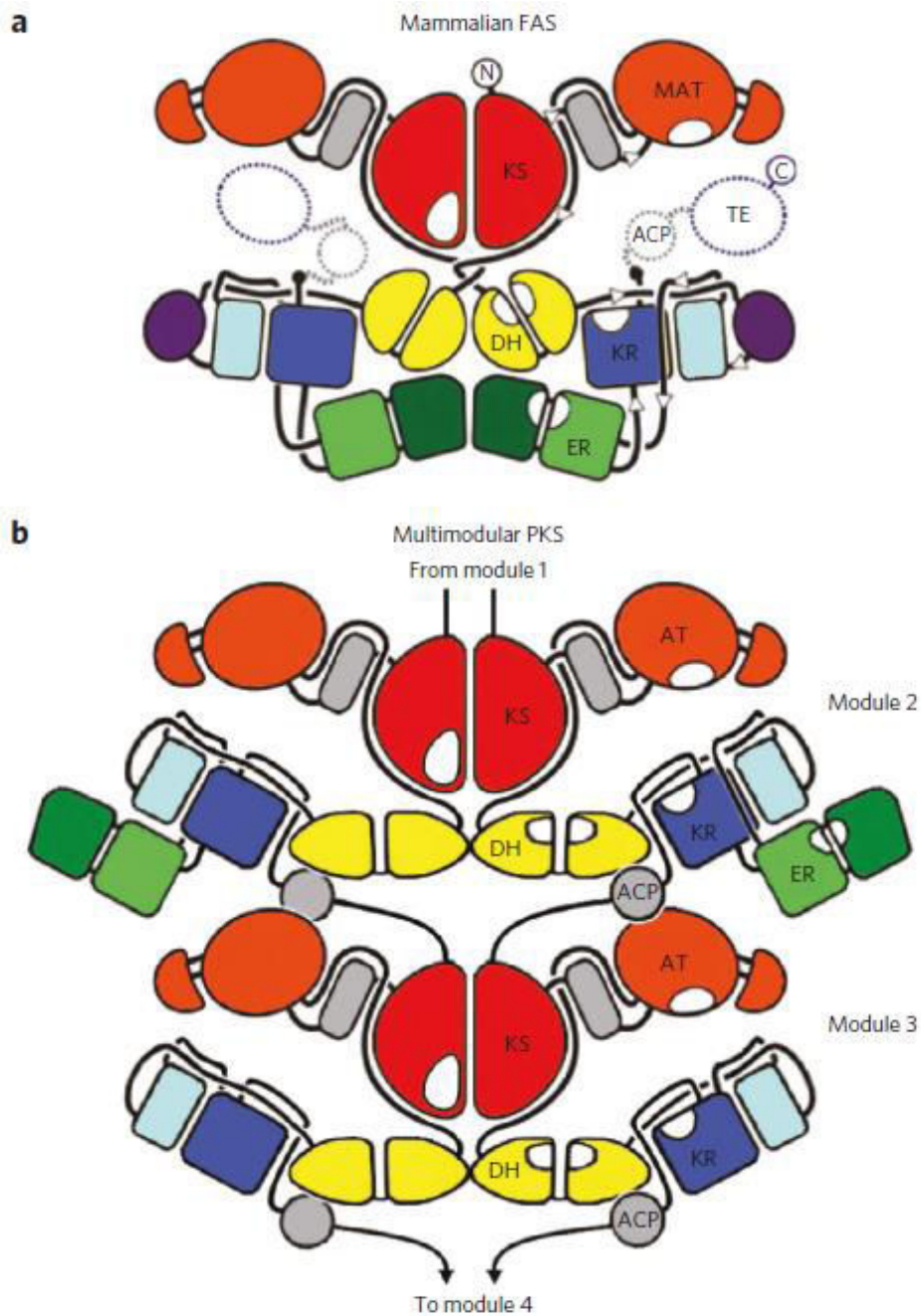


Figure 37 : Modèle d'organisation PKS extrapolé à partir de la structure de la mFAS (tiré de Zheng, et al., 2012).

(a) La structure de la mFAS se dimérise par l'intermédiaire des domaines KS, DH et ER. La position de l'ACP et du TE, bien que non attribuée de façon certaine, doit être à l'extrémité du linker en C-ter du domaine KR.

(b) Les structures des domaines et didomains PKS ont permis d'établir un modèle d'organisation pour les PKS de type I modulaires dans lequel le domaine ER serait monomérique et rejeté vers l'extérieur de la chambre réactionnelle. La présence d'une ER dimérique chez les PKS pourrait en effet empêcher l'ACP d'interagir avec le KS du module suivant.

2. Structure cristallographique d'une PKS itérative

La structure d'une PKS itérative responsable de la synthèse de l'acide micocérosique a été récemment résolue (Herbst, *et al.*, 2016). Cette PKS présente un mode de fonctionnement itératif, analogue aux FAS mais présente une forte homologie de séquence avec les PKS modulaires.

La structure a en fait été résolue en deux parties distinctes. Tout d'abord, le didomaine KS-AT a été résolu à 2,2 Å de résolution. Celui-ci présente un repliement similaire à ceux résolus auparavant (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007 ; Bretschneider, *et al.*, 2013), à l'exception près que la structure du KS-AT de la synthèse de l'acide micocérosique (MAS) révèle un monomère, et non un dimère. Bien qu'un équilibre monomère-dimère existe en solution, un K_d de 0,4 mM a été déterminé par ultracentrifugation (Herbst, *et al.*, 2016).

La structure de la région de modification, comportant les domaines DH, ΨKR, KR et ER, a été résolue à 3,75 Å de résolution. Dix-huit chaînes polypeptidiques sont présentes dans l'unité asymétrique. Tout comme les PKS modulaire, et par opposition aux FAS, aucun domaine ΨME n'est présent chez la MAS. Comme chez la mFAS, le domaine ER est dimérique, et participe, avec le domaine DH, à la dimérisation de la région de modification.

Un modèle regroupant les deux sous-ensembles a alors été proposé (Figure 38). Celui-ci est très similaire à la structure de la mFAS. En effet, tout comme la mFAS, deux chambres réactionnelles sont formées. Cependant, il s'agit là encore de la structure d'une synthase itérative, et bien que cette structure soit proposée comme représentative des PKS modulaires, un certain nombre de questions restent en suspens. En effet, les auteurs ont résolu la structure de la portion de condensation et de modification séparément. Or, cela soulève la question de la relevance de chacune des structures au sein de l'ensemble de la PKS.

De plus, chez la MAS, le domaine ER comporte un motif de dimérisation plus long que chez la majorité des PKS modulaires (Keatinge-Clay, 2016). De plus, pour les modules n'étant pas situés aux extrémités C-terminales des sous-unités, la modularité entraîne un lien covalent entre l'ACP et le KS du module en aval. Le linker, d'environ 15 à 20 résidus reliant ces deux domaines seraient alors trop court pour permettre à

l'ACP d'interagir avec tous ses partenaires (KS, AT, KR, DH, ER du même module et KS du module suivant) (Keatinge-Clay, 2016).

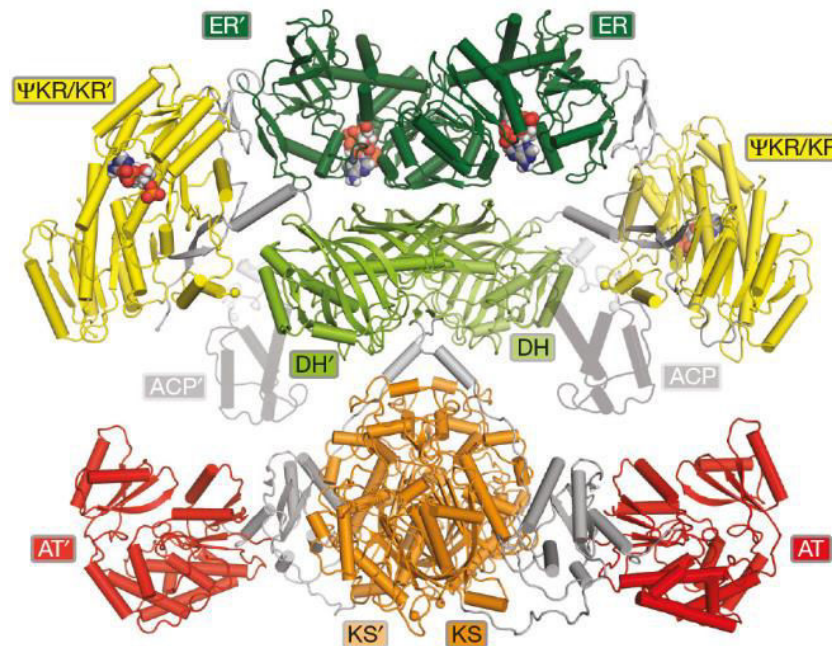


Figure 38 : Modèle proposé sur la base des structures du KS-AT et de la région de modification de la MAS (Herbst, et al., 2016).

3. Etat des connaissances sur les réarrangements conformationnels associés à la catalyse

En parallèle de mes travaux de thèse, deux études ont eu permis de résoudre la structure à faible ou moyenne résolution de modules entiers de PKS (Dutta, *et al.*, 2014 ; Edwards, *et al.*, 2014). Dans cette partie seront présentées ces études parues au même moment que notre publication, détaillées dans la partie résultat.

A) Structure d'un module entier par cryo-microscopie électronique (cryo-EM)

La structure cryo-EM de la sous unité PikAIII de la PKS synthétisant la pikromycine chez *S. venezuelae* (Dutta, *et al.*, 2014) contredit le modèle de la PKS (Zheng, *et al.*, 2012). La chaîne polypeptidique PikAIII ne comporte qu'un seul module, le module 5 de la PKS, qui doit donc interagir en *trans* à la fois avec le module précédant et le module suivant. Ce module est composé des trois domaines nécessaires à la condensation (AT, KS, ACP), d'un domaine KR et de deux interfaces de docking permettant l'interaction avec les modules partenaires.

Plusieurs structures représentant différents stades de la catalyse ont été résolues entre 7 et 11 Å de résolution, ce qui leur a permis de positionner des structures à haute résolution de domaines homologues dans les cartes obtenues. L'ensemble est dimérique par l'intermédiaire du domaine KS et, contrairement à ce qui était proposé jusqu'alors, les domaines forment une arche. En effet, les domaines AT et KR sont positionnés consécutivement en dessous du KS dimérique. Cette organisation donne ainsi naissance à une seule chambre réactionnelle et non deux comme ce qui était proposé à partir de la structure de la mFAS. Cette conformation est également en contradiction avec les structures des di-domaines KS-AT des modules 3 et 5 de DEBS dans lesquelles les domaines AT flanquent directement le KS (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007). Le linker KS-AT doit donc avoir une conformation différente de celle adoptée dans les structures cristallographiques afin d'aboutir à cette organisation. L'ACP *holo* (modifié par l'ajout du bras phosphopantéthéine), qui peut être clairement identifié dans la carte EM, adopte deux positions : la première proche du KR et la deuxième proche de l'AT, les deux retrouvés dans les mêmes proportions (Figure 39).

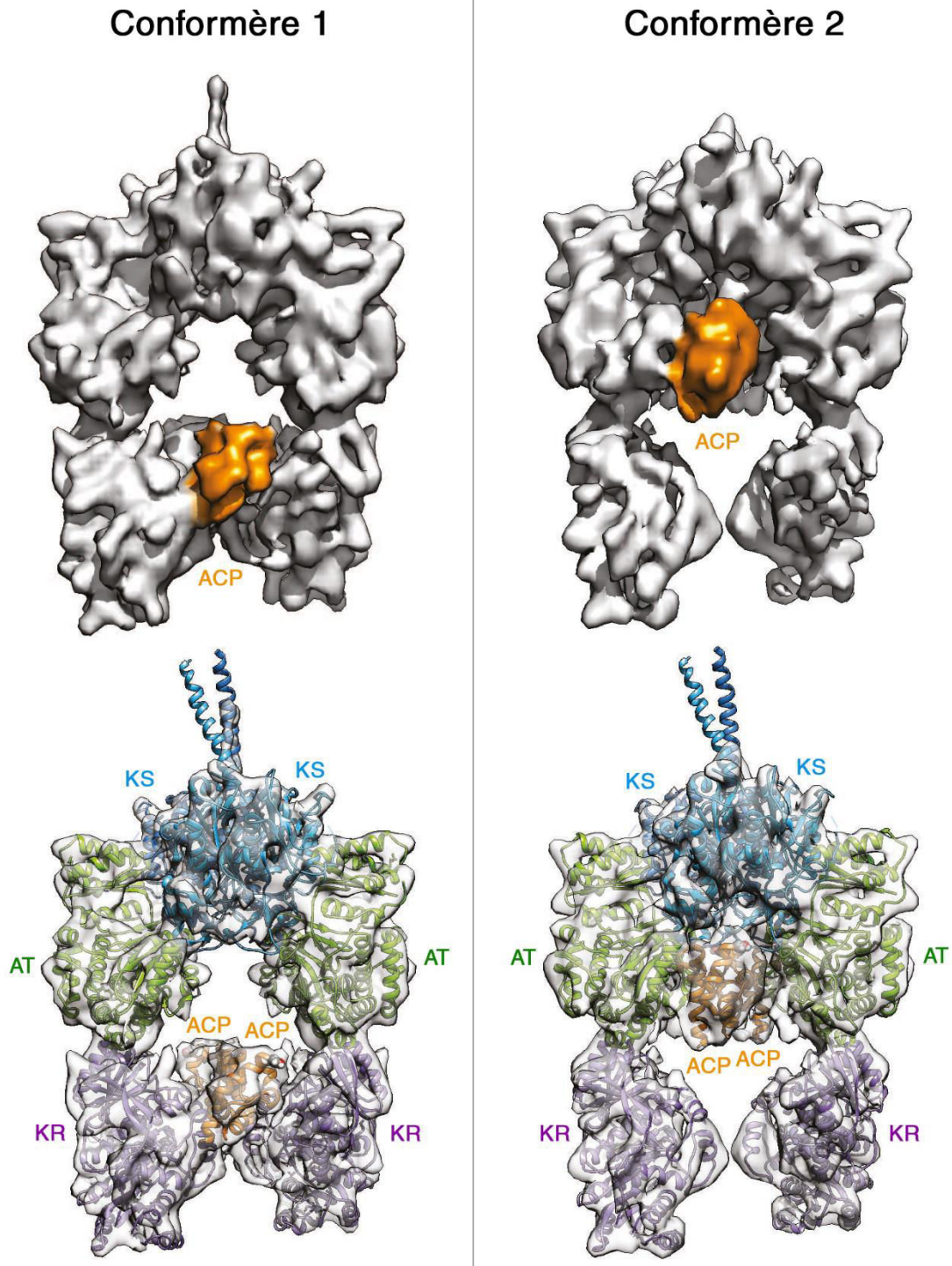


Figure 39 : Structure cryo-EM de PikAIII sous forme holo (adapté de Dutta, et al., 2014). Représentation de la structure cryo-EM de PikAIII en solide ou en transparent avec les structures à haute résolution positionnées au sein de la carte EM. L'ensemble s'agence en forme d'arche, le domaine KS dimérique formant le dôme et les domaines AT et KR se positionnant consécutivement en dessous. L'ACP (en orange) se trouve en contact soit avec le domaine KR (en violet) (conformère 1) ou avec le domaine AT (en vert) (conformère 2).

Ils se sont ensuite intéressés aux changements conformationnels nécessaires à la catalyse (Whicher, *et al.*, 2014). Pour cela, ils ont modifié tour à tour chacun des domaines ou des combinaisons de domaines avec l'objectif de bloquer le module à différentes étapes de la catalyse et analyser les changements de conformation par cryo-EM. Tout d'abord, dans le but de comprendre comment le KS₅ du module discrimine entre l'ACP₅ avec qui il interagit en *cis*, et l'ACP₄ avec qui il doit interagir en *trans*, ils ont réalisé une protéine de fusion entre l'ACP₄ et PikAIII(Δ ACP₅) dans laquelle l'ACP₅ de PikAIII a été délété pour l'empêcher d'interférer dans l'interaction du KS₅ avec l'ACP du module précédent. Ils ont ensuite modifié l'ACP₄ avec un pentakétide, son ligand naturel, et ce en utilisant une PPT de spécificité large : la Svp (Du, *et al.*, 2001). La cystéine du domaine KS a également été mutée pour empêcher le transfert du pentakétide. Ainsi l'ACP₄ semble venir se docker contre le KS₅ du côté opposé à la chambre réactionnelle, ce qui permettrait au KS de discriminer les deux ACP (Figure 40, conformation 1-2). Il est à noter que lorsque l'ACP₄ n'est pas modifié avec le pentakétide, il n'interagit pas avec le KS₅ ce que suggère que la présence du substrat est nécessaire afin d'établir une interaction stable.

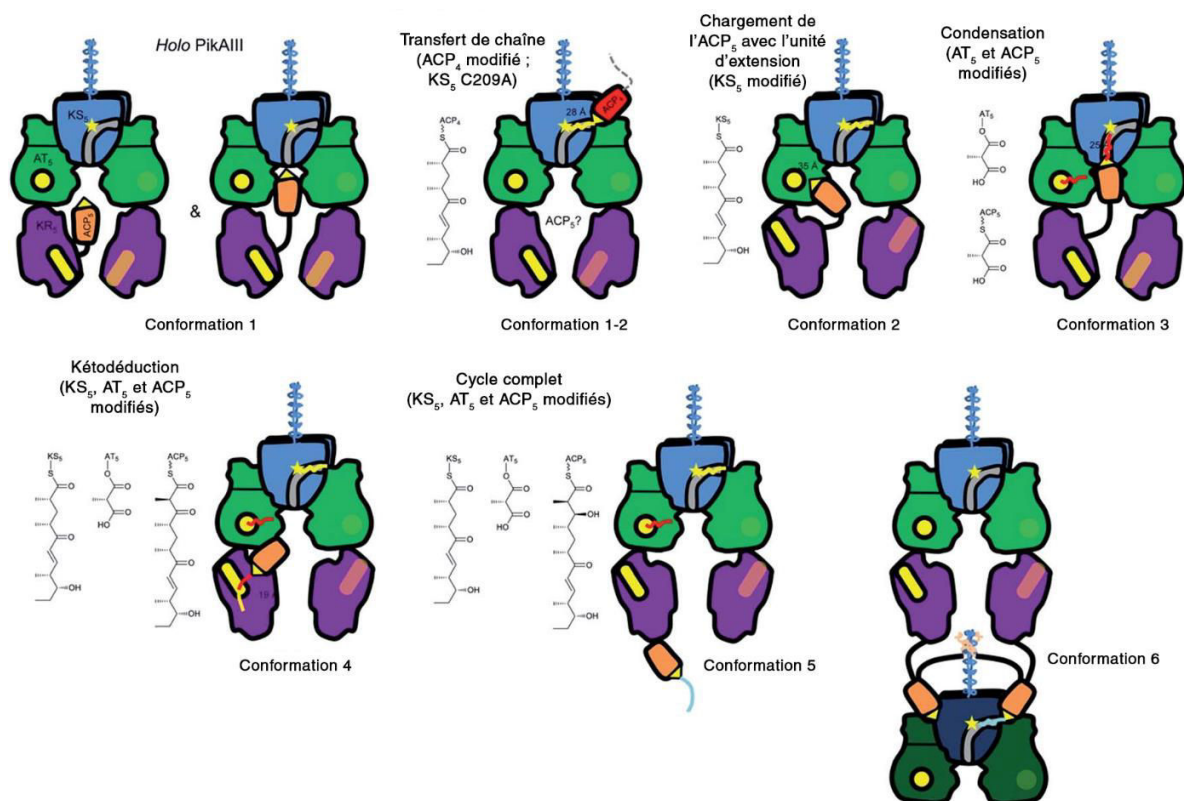


Figure 40 : Conformations adoptées par PikAIII au cours de la catalyse (adapté de Weissman, 2015).

Le domaine KS est représenté en bleu, l'AT en vert, la KR en violet et l'ACP en orange. Les molécules servant à la modification des domaines sont indiqués pour chaque conformation.

Ils ont ensuite acétylé spécifiquement le domaine KS₅ mimant ainsi la présence du substrat dans le site actif. Dans cet état, l'ACP₅ vient interagir avec le domaine AT₅. Cet état où le KS₅ est acylé avec la chaîne en extension et non l'AT₅ et l'ACP₅ avec des unités d'extension, n'est sans doute pas relevant d'un point de vue physiologique, et l'ACP₅ vient se placer à proximité de l'AT₅ pour être acylé. Ceci constituerait en quelque sorte un état d'urgence afin d'acyler rapidement l'ACP₅ (Weissman, 2015) (Figure 40, conformation 2). Il est également très étonnant de voir que la KR₅ opèrerait une rotation de presque 180° par rapport à la conformation précédente.

Afin de mimer la réaction de condensation, l'ACP₅ et l'AT₅ ont été modifiés. Le KS₅ n'a pas été acylé pour ne pas déclencher la réaction de condensation. L'ACP₅ se positionne alors juste en dessous du domaine KS₅, la sérine canonique de l'ACP₅ se trouvant à une vingtaine d'angströms du site actif du KS, distance qui peut être couverte par le bras phosphopantéthéine (Figure 40, conformation 3).

Ils sont ensuite parvenus à modifier l'ACP₅ avec un hexakétide sous forme cétone au niveau du carbone β. Ceci mime le polykétide après l'étape de condensation qui doit alors subir une réduction par le domaine KR₅. Il est à noter que leur protocole de modification de l'ACP₅ induit aussi la modification de l'AT₅ et du KS₅, ce qui n'est pas strictement relevant d'un point de vue physiologique (Weissman, 2015). L'ACP₅ ainsi modifié va venir interagir avec le KR₅, en se fixant entre l'AT₅ et le KR₅, et plaçant ainsi la sérine de l'ACP₅ à 19 Å du site actif du KR₅ qui est à nouveau dans une conformation retournée (Figure 40, conformation 4).

En incubant la construction précédente avec du NADPH, les auteurs ont pu obtenir le produit final du module. Une fois la cétone réduite par le domaine KR₅, l'ACP₅ se place directement sous le KR₅ afin de faciliter l'échange de la chaîne en élongation avec le module suivant (Figure 40, conformations 5 et 6).

B) Organisation structurale d'un module et bi-module par SAXS

La deuxième étude avait pour but de résoudre dans un premier temps la structure à faible résolution par SAXS du module 3 de DEBS fusionné au domaine TE. Cette construction permet en effet la synthèse d'un trikétide. Cela permet alors d'étudier les réarrangements conformationnels durant la catalyse en utilisant une construction de taille modérée. Ils ont ensuite résolu par la structure à faible résolution du bi-module 5 et 6 constituant la sous-unité DEBS 3 (Edwards, *et al.*, 2014).

Leur stratégie a tout d'abord consisté à analyser par SAXS les domaines constituant ces modules pour lesquels des structures à haute résolution étaient disponibles. Ils ont donc enregistré des données SAXS pour le di-domaine KS-AT, l'ACP₂, la KR₁ et le domaine TE. Cela a permis de mettre en évidence que les structures à haute résolution sont en adéquation avec les données SAXS. Ils ont ensuite analysé une construction dans laquelle le domaine TE est fusionné en C-terminale de l'ACP du module 3, une construction nommée M3+TE. Par modélisation en corps rigide, ils sont parvenus à proposer trois modèles rendant compte des données SAXS suggérant l'existence de plusieurs conformations en solution (Figure 41).

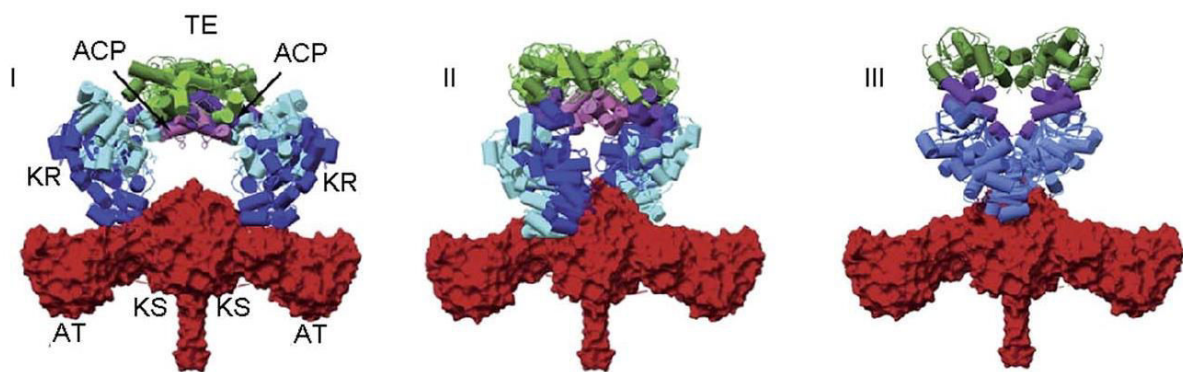


Figure 41 : Modélisation par corps rigide à partir des données SAXS du module 3+TE de DEBS (tiré de Weissman, 2015).

Le di-domaine KS-AT (en rouge) a été traité en tant que corps rigide lors de la modélisation, de même que le pseudo-dimère de KR (en bleu). L'ACP (en violet) se positionne à proximité du domaine TE (en vert), et ce, dans les trois conformations.

Cette organisation est très différente de celle obtenue lors des études par cryo-EM. Cela peut être dû au fait que lors de la modélisation, le KS-AT a été considéré comme un corps rigide, cela expliquant pourquoi l'AT est positionnée à l'extrémité de

la chambre réactionnelle (Weissman, 2015). L'ACP est d'ailleurs localisé à distance des sites actifs des domaines KR, AT et KS, et ne peut de cette façon interagir qu'avec le domaine TE. De plus, cette différence peut aussi être une conséquence de la troncature des données SAXS à de faibles valeurs de q ($q=0.15 \text{ \AA}^{-1}$) ce qui limite la résolution à environ 40 Å.

Les auteurs ont également résolu la structure à faible résolution de DEBS 3, composé des modules 5, 6 et du domaine TE. La position des domaines semble identique à celle du M3+TE, mais les deux modules sont tournés l'un par rapport à l'autre selon un angle de 70°.

L'un des objectifs de ma thèse était de résoudre la structure à faible résolution d'un module de PKS de type *trans*-AT. Le module 5 de la PKS synthétisant la virginiamycine (Figure 16) constituait l'objet de cette étude. Ce module est composé de trois domaines (un KS et deux ACP), et constitue un bon modèle pour comprendre les interactions entre les ACP et le domaine KS. Les caractéristiques du module 5 et les raisons de son choix comme modèle d'étude seront développés dans la partie IX.1.

En plus de comprendre les interactions entre domaines au sein d'un module, il est également primordial d'étudier comment s'effectue la communication entre les différentes sous-unités constituant les PKS.

VII. Les domaines de docking : gardiens de la continuité et de l'ordonnement de la synthèse des polykétides

Les interactions entre les différents modules composant la PKS sont capitales pour assurer la synthèse complète du polykétide. En effet, le polykétide en cours de synthèse doit être transféré d'un module à l'autre, passant alors du domaine ACP au domaine KS du module suivant. Cet échange implique uniquement ces deux partenaires lorsqu'il se fait en *cis* (donc lorsque les modules sont portés par la même sous-unité), mais devient plus complexe lorsqu'il s'effectue entre deux sous-unités. Dans ce cas, les interactions doivent être spécifiques afin d'empêcher le transfert de

la chaîne polykétidique vers une sous-unité portant les modules antérieurs, ou même avec une sous-unité provenant d'un autre système PKS au sein de la cellule. Ces interactions sont médiées, au moins en partie, par de courtes séquences peptidiques situées en N et C-terminales des sous-unités PKS appelés domaines de docking (DD).

Ces interfaces présentent un grand intérêt pour l'approche de biologie synthétique (Tang, *et al.*, 2000 ; Kittendorf, et Sherman, 2006 ; Menzella, et Reeves, 2007 ; Menzella, *et al.*, 2007). Comprendre les bases moléculaires de l'interaction entre DD permettrait en effet de pouvoir échanger des sous-unités d'une PKS par des sous-unités provenant d'une autre PKS. D'ailleurs, une étude visant à incorporer chez DEBS un module provenant d'une autre sous PKS a été réalisée (Gokhale, *et al.*, 1999). Dans cette étude, le module 2 de DEBS a été remplacé par le module 5 de la PKS synthétisant la rifamycine qui dispose de la même composition en domaine que le module 2 de DEBS. Le linker naturellement présent entre le module 1 et le module 2, de même que le DD en C-terminale du module 2 ont été conservés. Cette PKS hybride a produit le composé escompté sans perte de rendement (Figure 42).

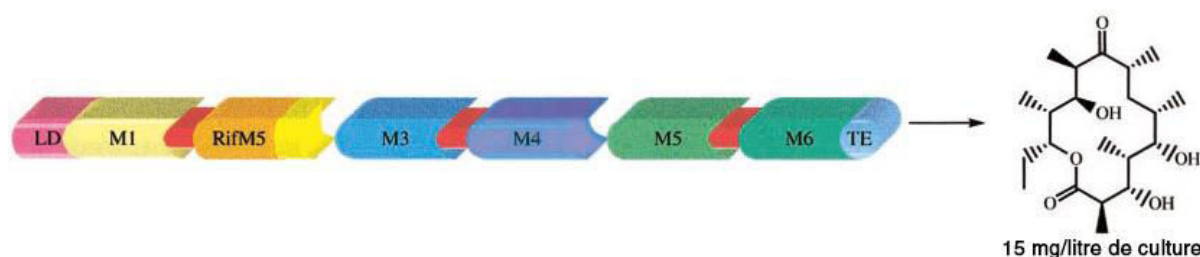


Figure 42 : PKS hybride dans laquelle le module 2 de DEBS a été remplacé par le module 5 de la PKS synthétisant la rifamycine (adapté de Gokhale, *et al.*, 1999)

Dans cette construction, le module 5 de la PKS synthétisant la rifamycine (en orange) est inséré en conservant le linker entre le module 1 et 2 (en rouge) et le DD entre DEBS 1 et DEBS 2 (en jaune).

L'implication des DD dans le transfert de la chaîne polykétidique d'une sous-unité à l'autre a ensuite été confirmée par une approche visant à échanger des DD chez DEBS, créant ainsi des paires non natives (Tsuji, *et al.*, 2001). Ainsi ils ont testé l'habileté de la PKS à transférer le polykétide du module 2 au module 3 (de DEBS 1 à DEBS 2) avec différent couple de DD. Lorsqu'ils étudient ce transfert pour les constructions comportant le DD en C-terminale du module 2 avec le DD en N-terminale du module 4 ou le DD en C-terminale du module 4 avec le DD en N-terminale du

module 3, le transfert reste possible mais est extrêmement ralenti. En revanche, lorsqu'ils reconstituent une paire de DD natifs en employant ceux en C-terminale du module 4 et en N-terminale du module 5, le transfert du polykétide se fait à une vitesse comparable à celle de l'interface normale. Ceci démontre que, même si ce n'est pas l'unique déterminant du transfert de la chaîne polykétidique d'une sous-unité à l'autre – les domaines ACP et KS sont aussi impliqués (Wu, *et al.*, 2002) – les DD ont une importance capitale dans l'interaction des sous-unités PKS entre elles.

La caractérisation biophysique d'un couple de DD s'est révélée délicate du fait de la faible interaction entre ces domaines. En effet, la constante de dissociation rendant compte de l'affinité de ces domaines est de l'ordre de la dizaine de micro molaires (Whicher, *et al.*). Ces interactions seraient transitoires au sein de la cellule, ce qui serait cohérent avec le fait que les sous-unités PKS ne sont pas liées entre elles mais interagissent de façon transitoire pour promouvoir le transfert du polykétide en cours d'élongation. Ceci permettrait également le renouvellement de sous-unités PKS défectueuses sans avoir recours à la néo-synthèse totale d'une nouvelle PKS, ce qui coûterait beaucoup d'énergie à la cellule.

La caractérisation structurale de couples de DD, ainsi que l'homologie de séquences, ont permis à ce jour de caractériser trois classes de DD qui seront décrites dans les chapitres suivants.

1. Les domaines de docking de type 1a

A cause de la faible interaction des DD, des protéines de fusion ont dû être produites afin de pouvoir résoudre la structure d'un couple de DD en interaction. Il a été démontré que la fusion des extrémités N et C-terminales des sous-unités PKS n'altéraient pas l'activité des PKS mais en plus augmentaient la stabilité des protéines (Squire, *et al.*, 2003 ; Liang, *et al.*, 1993)

Le premier couple de DD caractérisé est celui permettant l'interaction entre les sous-unités DEBS 1 et DEBS 2, ainsi composé du DD en C-terminale de l'ACP du module 2 et du DD en N-terminale du KS du module 3 (Broadhurst, *et al.*, 2003). Ces deux DD ont été fusionnés avec un linker de deux résidus et la structure de cette protéine de fusion a été résolue par RMN. Le DD en C-terminale de l'ACP est composé de 80 résidus alors que celui en N-terminale du KS comporte 37 acides aminés. Des

alignements de séquences, ainsi que des prédictions de structures secondaires, ont permis de mettre en évidence de nombreuses interfaces de docking similaires, qui appartiendraient donc au type 1a (Yadav, *et al.*, 2009).

La structure globale comporte deux domaines A et B. Le domaine A est formé des hélices 1 et 2 en C-terminale de l'ACP. Ce domaine A se dimérise, l'hélice 1 interagissant avec l'hélice 1' de l'autre monomère et les hélices 2 interagissant également entre elles. Par analogie avec le facteur de transcription HNF-1 qui possède un domaine présentant le même repliement et qui est impliqué dans la dimérisation, le domaine A est appelé domaine de dimérisation. Les interactions entre les hélices impliquent principalement des résidus hydrophobes.

Le domaine B est séparé du domaine A par une longue boucle d'environ 10 acides aminés. Il se compose de l'hélice 3 du DD en C-terminale de l'ACP et de l'hélice 4 du DD en N-terminale du KS. La longue hélice 4, constituée de 26 résidus, se dimérise en formant un coiled-coil. De chaque côté de ce coiled-coil viennent interagir les hélices 3 et 3' du DD en C-terminale de l'ACP. L'ensemble forme donc un fagot de 4 hélices α qui interagissent par l'intermédiaire d'interactions de type hydrophobes mais aussi par des ponts salins qui vont entraîner une spécificité d'interaction (Figure 43) (Broadhurst, *et al.*, 2003).

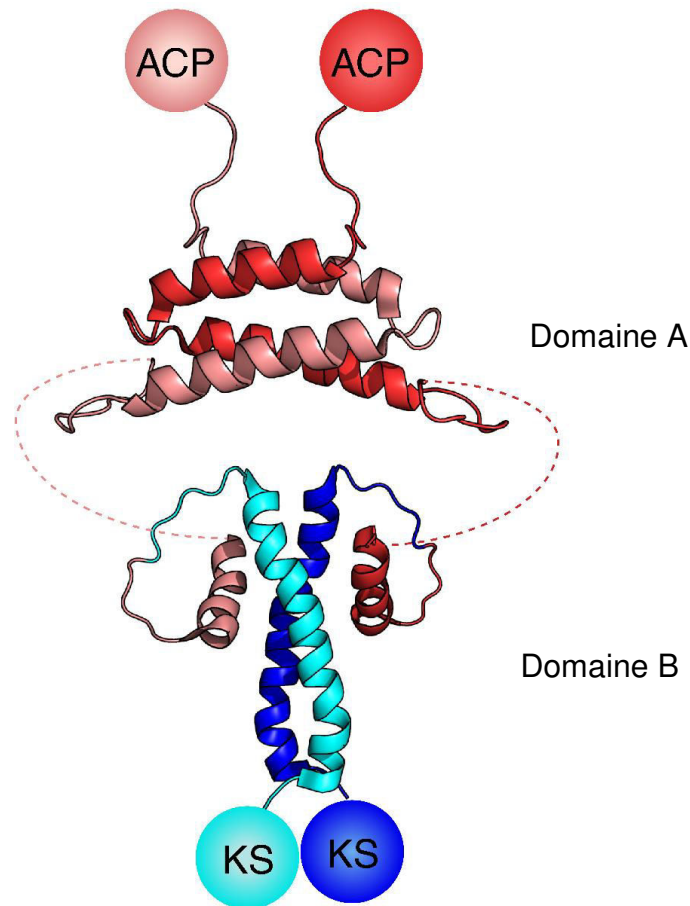


Figure 43 : Structure RMN des domaines de docking entre DEBS 1 et DEBS 2 (PDB : 1PZQ et 1PZR ; Broadhurst, et al., 2003).

Le DD en C-terminale de l'ACP de DEBS 1 (en rouge et en saumon) est constitué de 3 hélices. Les hélices 1 et 2 forment le domaine de dimérisation ou domaine A alors que l'hélice 3 interagit avec l'hélice 4 formant le DD en N-terminale du KS de DEBS 2 (en bleu et en cyan) pour former l'interface de docking ou domaine B. La boucle reliant le domaine A au domaine B est représentée en pointillés.

2. Les domaines de docking de type 1b

Autrefois appelés DD de type 2, les DD de type 1b présente une organisation similaire aux DD de type 1a. En effet, ils sont également composés d'un domaine de dimérisation constitué de deux hélices, et d'un domaine de docking à proprement parlé.

La structure de ce type de DD a été résolue par cristallographie pour le couple de DD entre les sous-unités 2 et 3 de la PKS synthétisant la pikromycine : PikAII et PikAIII (Buchholz, et al., 2009). Les auteurs ont tenté de produire une protéine de fusion incluant à la fois le domaine de dimérisation et le domaine de docking. Cette construction n'a en revanche jamais pu être cristallisée, sans doute à cause de la

flexibilité du linker séparant les deux interfaces. Ils ont donc produit une construction plus courte comportant uniquement l'interface de docking (sans le domaine de dimérisation). Dans cette construction, le DD en C-terminale de l'ACP est directement fusionné au DD en C-terminale du KS, sans aucun acide aminé servant de linker. Ils sont parvenus à résoudre la structure cristallographique à 1,75 Å de résolution.

L'absence de linker entre les deux DD ne permet pas une flexibilité suffisante pour que les deux DD interagissent au sein de la même chaîne polypeptidique. Ils ont donc obtenu une structure dans laquelle le DD en C-terminale de l'ACP interagit avec le DD en N-terminale du KS d'une deuxième protéine de fusion. Comme pour les DD de type 1a, le DD en N-terminale du KS se présente sous la forme d'une longue hélice α formant un coiled-coil. Le DD en C-terminale de l'ACP présente une hélice plus courte que celle des DD de type 1a. Cependant, une courte hélice supplémentaire est présente et vient se placer dans la jonction du coiled-coil (Figure 44).

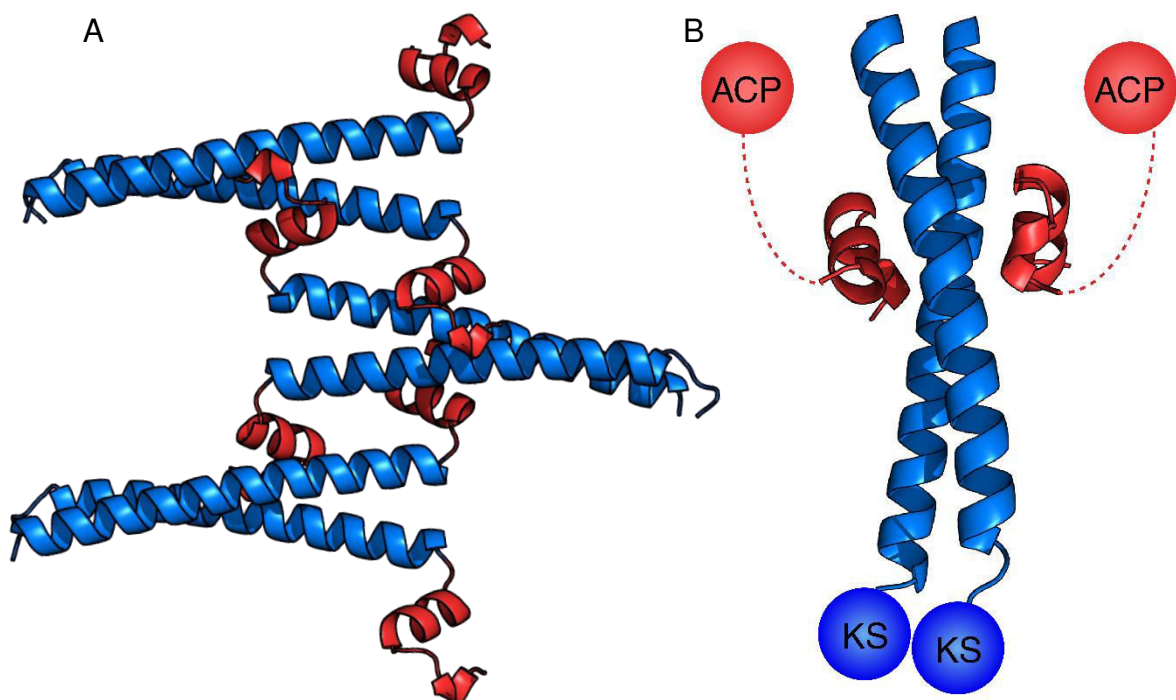


Figure 44 : Structure cristallographique de la fusion des DD entre PikAII et PikAIII (PDB : 3F5H ; Buchholz, et al., 2009).

A : Les interactions entre le DD en C-terminale de PikAII (en rouge) et le DD en N-terminale de PikAIII (en bleu) impliquent trois dimères différents, la boucle séparant les deux DD n'étant pas assez longue pour permettre l'interaction à travers la même chaîne peptidique.

B : Le DD en N-terminale du KS forme une longue hélice qui se dimérise en coiled-coil. Le DD en C-terminale de l'ACP est, quant à lui, formé d'une hélice et d'un tour d'hélice, les deux interagissant avec le coiled-coil *via* des interactions de type hydrophobe et des ponts salins.

3. Les domaines de docking de type 2

Grâce à des analyses de séquences, les DD de la PKS synthétisant la curacine semble présenter certaines particularités par rapport aux DD de type 1 (Whicher, *et al.*). Il en va de même pour de nombreux autres systèmes PKS, suggérant l'existence d'une classe supplémentaire de DD. En effet cette analyse révèle que le DD en C-terminale de l'ACP comporterait environ 40 acides aminés de moins que ceux des DD de type 1 et serait donc trop court pour inclure un domaine de dimérisation. Le DD en N-terminale du KS, bien que de la même longueur que les DD de type 1, serait trop polaire pour former un seul long coiled-coil.

Les structures de deux couples de DD de la PKS synthétisant la curacine (entre les sous unités CurG/CurH et CurK/CurL) ont été résolues par cristallographie. Les DD ont été fusionnés entre eux par l'intermédiaire d'une linker flexible de 8 résidus : G₃SG₃S permettant des interactions natives entre les DD.

Concernant le couple de DD entre CurG et CurH, chaque DD se structure en deux hélices α reliées par une courte boucle. La deuxième hélice du DD en N-terminale du KS se dimérise en formant un coiled-coil. En revanche, la première hélice, plus polaire, se retrouve de chaque côté du coiled-coil dimérique. Le DD en C-terminale de l'ACP est également composé de deux hélices α qui interagissent chacune avec une hélice du DD en N-terminale du KS (Figure 45). L'ensemble forme donc un fagot de 8 hélices α . Bien que l'affinité soit comparable à celle des DD de type 1, cette organisation permet d'augmenter la surface d'interaction entre les DD. Ceci est cohérent avec le fait que la PKS synthétisant la curacine est composé uniquement de modules discrets qui doivent donc interagir en *trans*. Il est donc nécessaire que ces interactions soient très spécifiques. Un autre point intéressant concerne la topologie de cette interaction. En effet ce mode d'interaction entre les deux DD permet de positionner l'ACP directement à côté du KS, alors que pour les DD de type 1, les linker reliant le DD à l'ACP et celui reliant l'autre DD au KS sont à l'opposé.

Le couple de DD entre CurK et CurL est légèrement différent. En effet, le DD en N-terminale du KS n'est pas composé de deux hélices mais d'une seule hélice, plus longue, qui se dimérise également en formant un coiled-coil. De la même façon, le DD

en C-terminale de l'ACP est constitué d'une hélice plus longue que celle de CurG et également d'une très courte hélice (Figure 45) (Whicher, *et al.*).

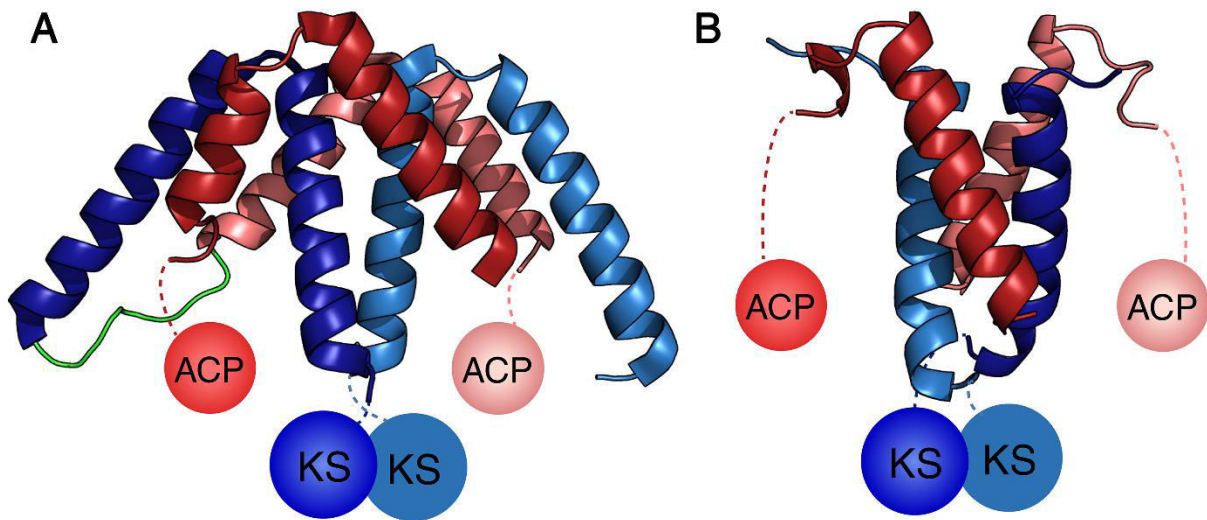


Figure 45 : Structures des DD de type 2 mis en évidence chez la PKS synthétisant la curacine (PDB : 4MYY et 4MYZ ; Whicher, *et al.*, 2013).

A : Structure de la fusion entre les DD de CurF (en rouge) et de CurG (en bleu). Chaque DD est constitué de 2 hélices et l'ensemble est dimérique, ce qui forme un fagot de 8 hélices α .

B : Structure de la fusion entre les DD de CurK (en rouge) et de CurL (en bleu). Le DD de CurK est formé d'une longue hélice ainsi que d'un tour d'hélice. Le DD de CurL est quant à lui composé d'une seule hélice α . Aucun de ces DD ne présente de domaine de dimérisation comme c'est le cas pour les DD de type 1.

4. Une classe à part de domaines de docking

Une autre classe de DD a été caractérisée en 2008 chez le système hybride PKS/NRPS responsable de la synthèse de tubulysine (Richter, *et al.*, 2008). Ce couple de DD est situé à l'interface entre deux modules NRPS, entre les sous-unités TubB et TubC.

La structure du DD en N-terminale de TubC (nommé TubCdd) a été résolue par RMN. Ce DD de 70 résidus se structure en trois hélices α et deux brins β qui adoptent une topologie $\alpha\beta\beta\alpha$. Ce DD s'homodimérise par l'intermédiaire des deux brins β , mais également de la boucle reliant le deuxième brin à l'hélice 2 (Figure 46).

Ils ont ensuite cherché à résoudre la structure des deux DD en interactions, et ont pour cela tenté de produire une protéine de fusion dans laquelle les deux DD ont été fusionnés par un linker de deux acides aminés. Cependant une analyse de la protéine

par spectrométrie de masse a permis de mettre en évidence une protéolyse au niveau du N-terminale de la protéine de fusion (TubBdd). Ils ont donc cherché à produire le TubBdd par synthèse chimique – il s'agit d'un peptide de 25 résidus – mais celui-ci s'est révélé non structuré par RMN et dichroïsme circulaire. Cependant, ce dernier a tout de même pu être utilisé pour tester l'implication dans l'interaction des résidus chargés à la surface de TubCdd.

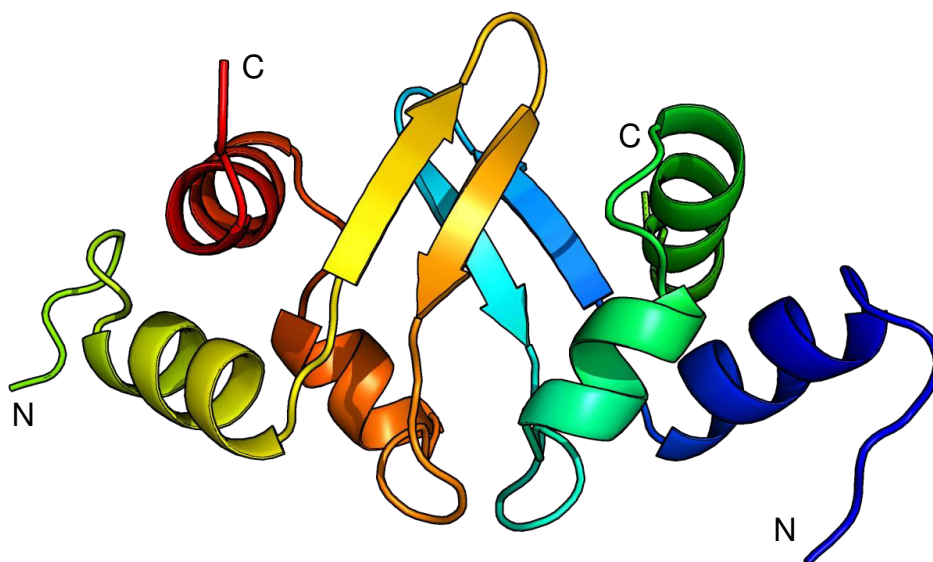


Figure 46 : Structure RMN du dimère de TubCdd (PDB : 2JUG, adapté de Richter, et al., 2008).

Chaque monomère se structure en trois hélices α et un feuillet β composé de deux brins. L'ensemble se dimérise par l'intermédiaire de ponts salins au niveau des feuillets et dans la boucle reliant le feuillet à l'hélice 2.

L'un des aspects de ma thèse consistait à caractériser au niveau structural et biochimique, l'interaction entre les DD en C-terminale du module 5 et N-terminale du module 6 de la PKS synthétisant la virginiamycine M.

VIII. Les enzymes de maturation post-PKS : une cible de choix pour l'ingénierie ?

Généralement, suite à la synthèse du polykétide par la PKS, celui-ci est mûré par des enzymes qui vont le modifier chimiquement pour lui conférer sa structure finale et son activité biologique. Les gènes codant ces enzymes post-PKS sont localisés au sein du cluster de gène codant la PKS. Au vu de leur importance cruciale pour la bio-activité des polykétides, les enzymes post-PKS sont des cibles de choix pour la biologie synthétique. Ceci peut se faire par la délétion d'une enzyme post-PKS, l'ajout d'une activité supplémentaire, ou bien encore la modification rationnelle d'une de ces enzymes pour altérer la stéréospécificité, par exemple (Weissman, 2016).

On compte parmi ces enzymes de nombreuses oxyréductases. Ces enzymes sont responsables de l'ajout de groupement hydroxyle, aldéhyde ou cétones ou encore de la réduction d'une de ces fonctions. Ces groupements peuvent constituer des bases pour la modification ultérieure par d'autres enzymes post-PKS telles que les glucosyltransférases (Rix, *et al.*, 2002). Dans cette catégorie, des cytochromes P450 monooxygénases sont particulièrement attractifs pour l'approche d'ingénierie, puisqu'ils agissent dans les ultimes étapes de la biosynthèse des polykétides. De plus, on trouve des P450 de spécificité large, telle qu'EryK qui intervient dans la voie de synthèse de l'érythromycine A, ou bien encore PikC qui agit dans la synthèse de la pikromycine. Ces deux enzymes, capables d'oxyder un grand nombre de molécules, sont donc particulièrement utilisées dans les études d'ingénierie (Lee, *et al.*, 2004 ; Desai, *et al.*, 2004)

De nombreux polykétides sont également glycosylés. L'ajout de ces unités osidiques confère l'activité de certaines molécules comme l'érythromycine A ou l'amphotéricine B, puisque ces sucres participent à l'interaction avec les cibles de ces polykétides (Schlünzen, *et al.*, 2001 ; Wahab, *et al.*, 2008). Les unités osidiques ajoutées sont de natures diverses, appartenant dans la grande majorité des cas à la famille des 6-désoxyhexoses. Bien que des monosaccharides soient fréquemment ajoutés, il peut arriver que des sucres à 7 unités soient transférés au polykétide. L'ingénierie des glycosyltransférases peut se faire à travers la délétion de ces enzymes, avec ou sans complémentation par une autre glycosyltransférase, ou bien

par expression d'une cassette de gènes contenant tous les gènes nécessaires à la synthèse et à l'activation de sucres (Gaisser, *et al.*, 2002 ; Schell, *et al.*, 2008).

Parmi les enzymes de modification post-PKS on retrouve également des méthyltransférases, diverse transférases (acyltransférases, carbamoyltransférases, aminotransférases, etc.) ou bien encore des enzymes catalysant des réactions d'halogénéation. Soit autant de cibles potentielles pour l'ingénierie et la génération de nouveaux polykétides (Olano, *et al.*, 2010).

Bien que les domaines TE libèrent généralement le polykétide en catalysant sa macrocyclisation, quelques cyclases sont impliqués dans la dernière étape de synthèse du polykétide. C'est notamment le cas pour la lankacidine, libéré sous forme d'un précurseur linéaire par le domaine TE qui n'entraîne que la formation d'une lactone à 6 membres. C'est ensuite une amine oxydase qui est proposée pour catalyser la macrocyclisation du polykétide (voir chapitre IX.2). Dans d'autre cas, plusieurs oxygénases ont été décrites pour intervenir dans la cyclisation de polykétide. C'est notamment le cas de la rifamycine B, pour laquelle une hydroxylase catalyse la formation du noyau naphthalène, même si ici, l'enzyme agit entre la deuxième et la quatrième étape d'élongation (Xu, *et al.*, 2005). De même, un cytochrome P450 monooxygénase catalyse la formation d'un cycle tétrahydrofurane lors de la synthèse de l'auréothine, un antifongique et antibiotique, suite à sa libération par la PKS (He, *et al.*, 2004).

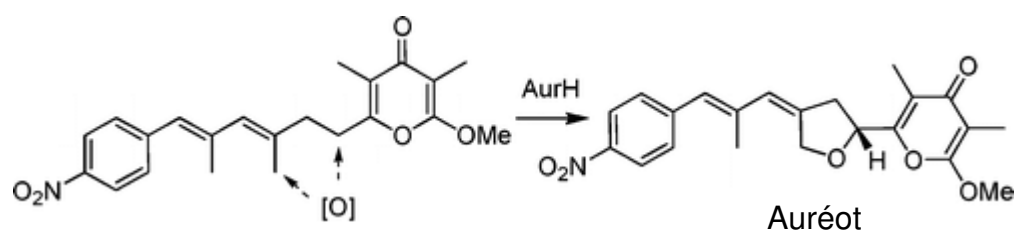


Figure 47 : Réaction de formation du cycle tétrahydrofurane catalysée par l'enzyme post-PKS AurH (tiré de He, *et al.*, 2004).

Ces enzymes représentent donc un large vivier pour l'approche de biologie synthétique. Leur utilisation est pertinente étant donné que ces enzymes post-PKS interviennent au stade final de la synthèse du polykétide et, de ce fait, une modification de ces enzymes n'entraînera pas de problème de spécificité de substrat (à l'exception

des enzymes post-PKS intervenant en aval de la modification), comme ça peut être le cas avec les domaines en aval d'une modification de la PKS.

L'un de mes objectifs durant ma thèse était de caractériser une enzyme post-PKS catalysant la cyclisation de la lankacidine, la LkcE.

IX. Choix de deux systèmes modèles de PKS de type *trans*-AT

Comme nous l'avons vu précédemment, les PKS de type *trans*-AT ont été très peu caractérisée jusqu'à présent. Comme cible de nos études, nous avons choisi deux système modèles de PKS de type *trans*-AT, qui sont à la fois représentative de ce type de système et qui sont responsables de la synthèse de molécules à fort potentiel pharmacologique.

1. La virginiamycine M et sa PKS

La virginiamycine M est un antibiotique de la famille des streptogramines. Ces derniers sont produits majoritairement par des bactéries du genre *Streptomyces*. Un dérivé hémisynthétique de la virginiamycine M (la dalfopristine) est utilisé pour traiter différentes infections par des bactéries Gram positif résistantes à la vancomycine. Elle est généralement utilisée avec un dérivé de la virginiamycine S (la quinupristine) avec qui elle présente un effet synergique. Les antibiotiques de la famille des streptogramines sont classés en deux catégories : le type A qui sont des hybrides polykétide-peptide cycliques et le type B qui sont des hexa ou heptapeptides cycliques, et les deux classes sont connues pour leur synergie (Johnston, *et al.*, 2002). En effet, individuellement, ces antibiotiques ont un effet bactériostatique, mais utilisé en combinaison, ils présentent une activité accrue de 100 fois (Cocito, 1969 ; Cocito, 1979). D'ailleurs un traitement à base de dalfopristine et quinupristine est commercialisé sous le nom de Synergid®.

La virginiamycine M est un inhibiteur de la synthèse protéique. Elle se fixe à la grande sous-unité ribosomique 50S des procaryotes, à cheval entre les sites P et A empêchant ainsi la fixation correcte des ARNt dans le site P (Nyssen, *et al.*, 1989 ; Hansen, *et al.*, 2003). En plus de l'inhibition partielle de cette sous-unité ribosomique,

la fixation de la virginiamycine M va entraîner un changement conformationnel du ribosome permettant à la virginiamycine S de se fixer, d'où son effet synergique.

Ces deux antibiotiques sont produits par des bactéries du genre *Streptomyces*, et notamment par *S. virginiae*. Alors que la virginiamycine S est synthétisée par un système NRPS (Namwat, *et al.*, 2002), le cluster de gènes responsable de la synthèse de la virginiamycine M a été séquencé et annoté, révélant des gènes codant une hybride PKS-NRPS (Pulsawat, *et al.*, 2007). La production de virginiamycine (M et S) est induite par la synthèse de γ -butyrolactone par *S. virginiae*. Cette dernière permet en effet de surexprimer les gènes codant les deux synthèses (Namwat, *et al.*, 2002). Au sein du cluster de gènes se trouve également des gènes codant des ABC transporteurs pour la résistance et des protéines régulatrices.

A ce jour, trois gènes codant les sous-unités de la PKS ont été identifiés. Celles-ci comportent en tout 9 modules, le gène codant la sous-unité portant le module 10 responsable de l'incorporation de la proline terminale n'ayant pas été trouvé lors du séquençage (Figure 49) Cependant le gène codant ce module a été identifié dans un cluster homologue chez *S. pristinaspiraelis* (Mast, *et al.*, 2011). De même, un module de chargement, absent dans l'annotation du cluster de gènes codant la PKS de la virginiamycine, a été découvert lors du séquençage du génome de *S. pristinaspiraelis*.

Le gène *virA* code une protéine hybride PKS-NRPS qui va contenir les cinq premiers modules responsables de cinq cycles d'élongation. Les deux premiers modules permettent l'incorporation d'unités malonate, le module 3 est un module NRPS permettant l'incorporation d'une glycine, et les deux modules suivants catalysent l'ajout de deux malonates supplémentaires.

Un constat notable est que les modules 1 et 5 contiennent deux domaines ACP. Le fait d'avoir deux ACP successifs pourrait avoir plusieurs rôles. Cela pourrait permettre d'augmenter la production du polykétide, les ACP agissant en parallèle (Rahman, *et al.*, 2005). Ou bien, L'un des avantages de ce duo d'ACP pourrait être de favoriser les interactions avec les domaines de modification en nécessitant ainsi des changements conformationnels moins importants. Il est aussi proposé que ces tandems d'ACP puissent également servir à maintenir suffisamment longtemps le polykétide au sein du module, avant le prochain cycle de catalyse, pour permettre à des réactions de modifications d'avoir lieu en *trans* (Calderone, *et al.*, 2007).

Justement, une cassette de β -méthylation intervient entre le module 5 et le module 6 pour catalyser le transfert d'un groupement méthyle sur le carbone β du polykétide. Ces cassettes de β -méthylation typique des PKS de type *trans*-AT, aussi appelées cassettes hydroxyméthylglutaryl-CoA-synthase (Cassette HMGS), sont généralement composées de cinq protéines (Calderone, *et al.*, 2006 ; Calderone, 2008). En effet, cette méthylation nécessite la présence d'une KS discrète, d'un ACP discret, d'une protéine homologue aux HMG-CoA synthase (HCS), d'une énoyl-CoA déshydratase (ECH1) et d'une crotonase (ECH2). Dans un premier temps, un malonyle attaché au bras phosphopantéthéine de l'ACP (VirAB – découvert dans le cluster de gènes par KJ. Weissman) est décarboxylé par la KS (VirB) dont les résidus catalysant les réactions de condensation sont absents. L'ACP transfère ensuite l'acétyle ainsi obtenu sur la cystéine catalytique de la HMG-CoA synthase (VirC). La HCS ainsi acylée est activée et va catalyser la réaction d'addition d'aldol sur le polykétide en extension (apporté par l'ACP_{5a} ou l'ACP_{5b}). L'hydroxyméthylglutaryle ainsi formé va ensuite subir une déshydratation catalysée par l'énoyl-CoA déshydratase (VirD), puis une décarboxylation catalysée par la crotonase (VirE) (Figure 48).

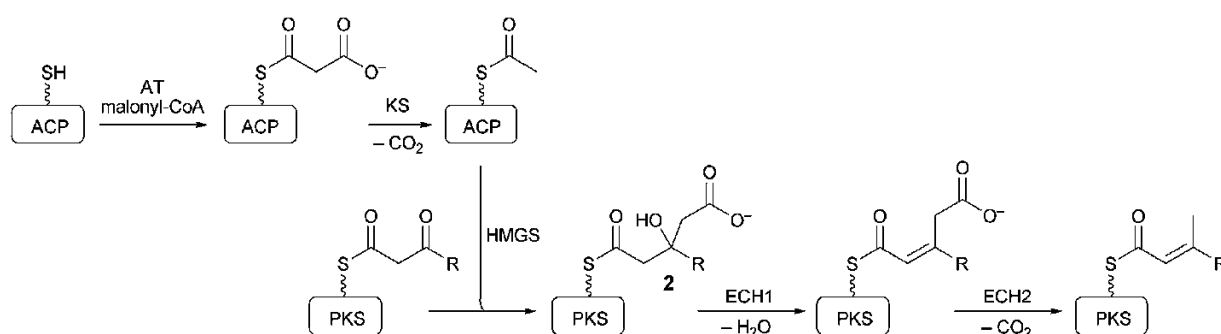


Figure 48 : Mécanisme de méthylation sur la position β des polykétides (tiré de Piel, 2010). Tout d'abord, une AT charge le malonate sur l'ACP. Celui-ci interagit avec la KS qui décarboxyle le malonate entraînant ainsi la formation d'un acétyl-ACP. Ce groupement est ensuite transféré au niveau de l'HMG-CoA synthase qui va catalyser son addition sur le polykétide en cours d'élongation au niveau du carbone β . S'ensuit une déshydratation catalysée par l'énoyl-CoA déshydratase (ECH1) puis une décarboxylation catalysée par la crotonase (ECH2).

Alors que le séquençage prédisait que le module 6 était scindé sur deux protéines, notre équipe a démontré que VirF et VirG ne forment en fait qu'une seule et même chaîne peptidique appelée VirFG. Cette dernière porte les modules 6 et 7 responsables de l'incorporation de deux unités malonates supplémentaires. La sous-unité VirH porte, quant à elle, le module NRPS responsable de l'incorporation d'une

sérine et un module PKS inactif composé d'un domaine KS^0 et d'un ACP. Ce module joue sans doute un rôle structural important, permettant aussi à VirH de conserver l'état oligomérique dimérique du reste de la PKS.

Aucun module PKS ne comporte de domaine AT, c'est donc la protéine discrète VirI qui se charge d'acyler chacun des ACP en *trans*. De même, l'activité thioestérase II (*proof-reading*) est portée par la protéine discrète VirJ (Figure 49).

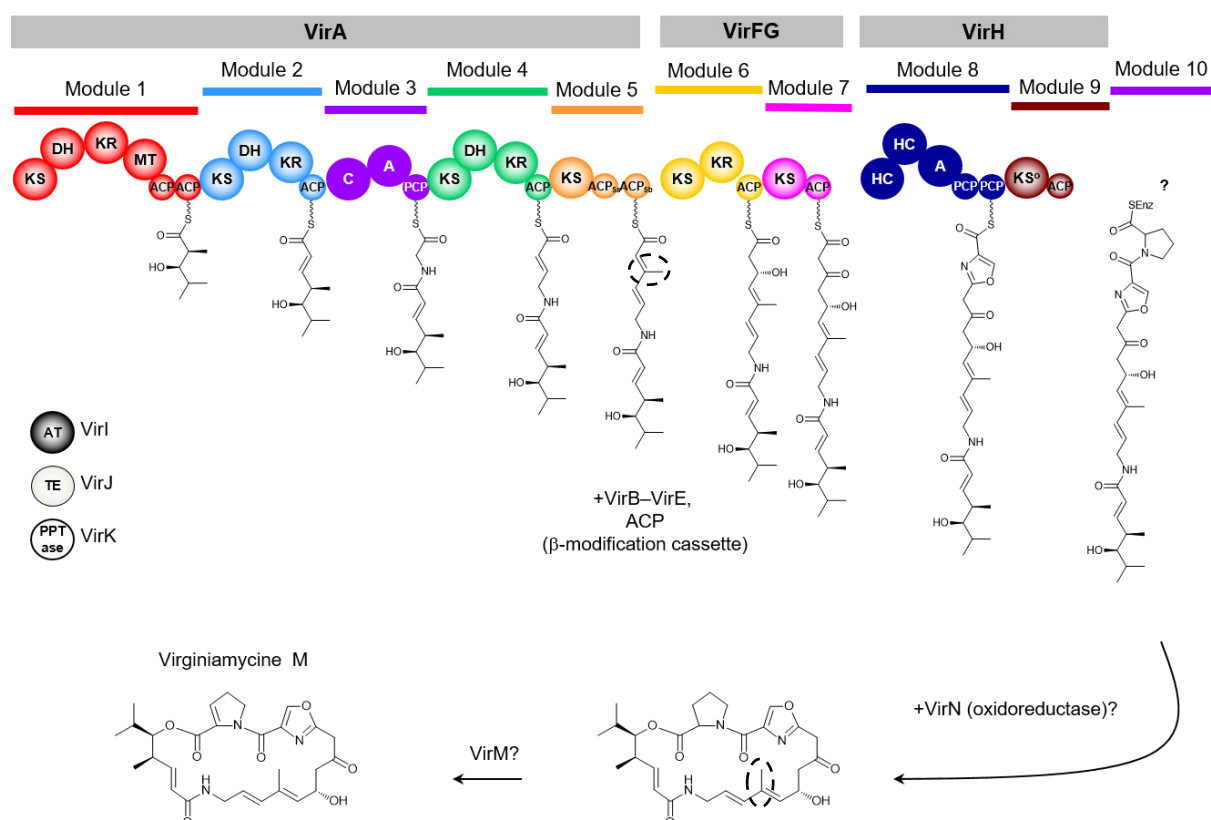


Figure 49 : Voie de synthèse proposée pour la virginiamycine M (tiré de Davison, et al., 2014).

La PKS complète comporte sept modules PKS et deux modules NRPS (modules 3 et 8). Le module de chargement et le module 10 sont manquants dans le séquençage de la PKS. Au niveau du module 5, une méthylation sur la position β est introduite par VirB–E et un ACP (VirAB). La position de cette méthylation est indiquée par un cercle en pointillé. VirI constitue l'AT discrète et VirJ une TEII. Le dernier module responsable de l'incorporation de la proline est manquant dans le séquençage de la PKS et n'a donc toujours pas été identifié. Le module 9 est inactif ; la KS ne possédant pas d'histidine catalytique est nommée KS^0 .

2. La PKS synthétisant la lankacidine C.

La lankacidine est un antibiotique qui est couramment utilisé en médecine vétérinaire pour traitement du bétail infecté par des bactéries Gram positif pathogènes (Hayashi, *et al.*, 1988). Elle est commercialisée sous le terme de sédécamycine ou bien son variant hémi-synthétique : la terdécamycine (Yatsukawa, *et al.*, 2008). La lankacidine semble également posséder un rôle d'anticancéreux contre les cellules leucémiques (Oostu, *et al.*, 1975).

Il s'agit d'un inhibiteur de la synthèse protéique, en se fixant sur le ribosome au niveau du centre peptidyl-transférase de la grande sous-unité ribosomique (Auerbach, *et al.*, 2010). Tout comme la virginiamycine M, la lankacidine agit en synergie avec un autre polykétide, la lankamycine. La production des deux antibiotiques est d'ailleurs co-régulée (et co-induite par la présence de γ -butyrolactone), suggérant que les deux polykétide ont été optimisés dans l'évolution pour agir simultanément (Arakawa, *et al.*, 2007 ; Yamamoto, *et al.*, 2008). La structure cristallographique de la sous-unité ribosomique 50S en complexe avec la lankacidine a été résolue, permettant d'identifier le site de fixation de chaque antibiotique (Auerbach, *et al.*, 2010 ; Belousoff, *et al.*, 2011). La lankacidine se lie au niveau du site A, empêchant ainsi le positionnement adéquat de l'acide aminé porté par l'ARNt. De plus, la fixation de la lankacidine n'altère en rien la fixation de la lankamycine, ce qui explique l'effet synergique.

Les deux systèmes PKS sont codés par un plasmide qui code également les gènes de régulation : le plasmide PSLA2-L retrouvé chez *S. rochei* (Kinashi, *et al.*, 1994). Ce plasmide de 200 kb a été séquencé, révélant la présence de quatre clusters de gènes responsables de la synthèse de quatre métabolites secondaires (Suwa, *et al.*, 2000).

La lankacidine est assemblée à partir de huit unités malonates et d'une glycine. Cependant, seules cinq KS sont présentes au sein du cluster, ce qui suggère que plusieurs modules agissent de façon itérative. Plusieurs voies de synthèse ont été proposées pour la lankacidine (Piel, 2010). La plus probable impliquerait la protéine LkcA, qui comporte le module de chargement NRPS ainsi qu'un domaine KS, et qui catalyserait l'incorporation de la glycine en tant qu'unité de départ. Ensuite, huit unités malonates supplémentaires sont ajoutées. Pour cela, la LkcC et la LkcF agirait tour à tour pour catalyser respectivement trois et deux cycles d'extension. Enfin, la protéine LkcG catalyserait le dernier cycle d'extension et la libération du polykétide. Il est à

noter qu'une DH discrète, LkcB intervient lors de plusieurs cycles d'extension (Figure 50) (Dickschat, *et al.*, 2011 ; Tatsuno, *et al.*, 2007 ; Tatsuno, *et al.*, 2009).

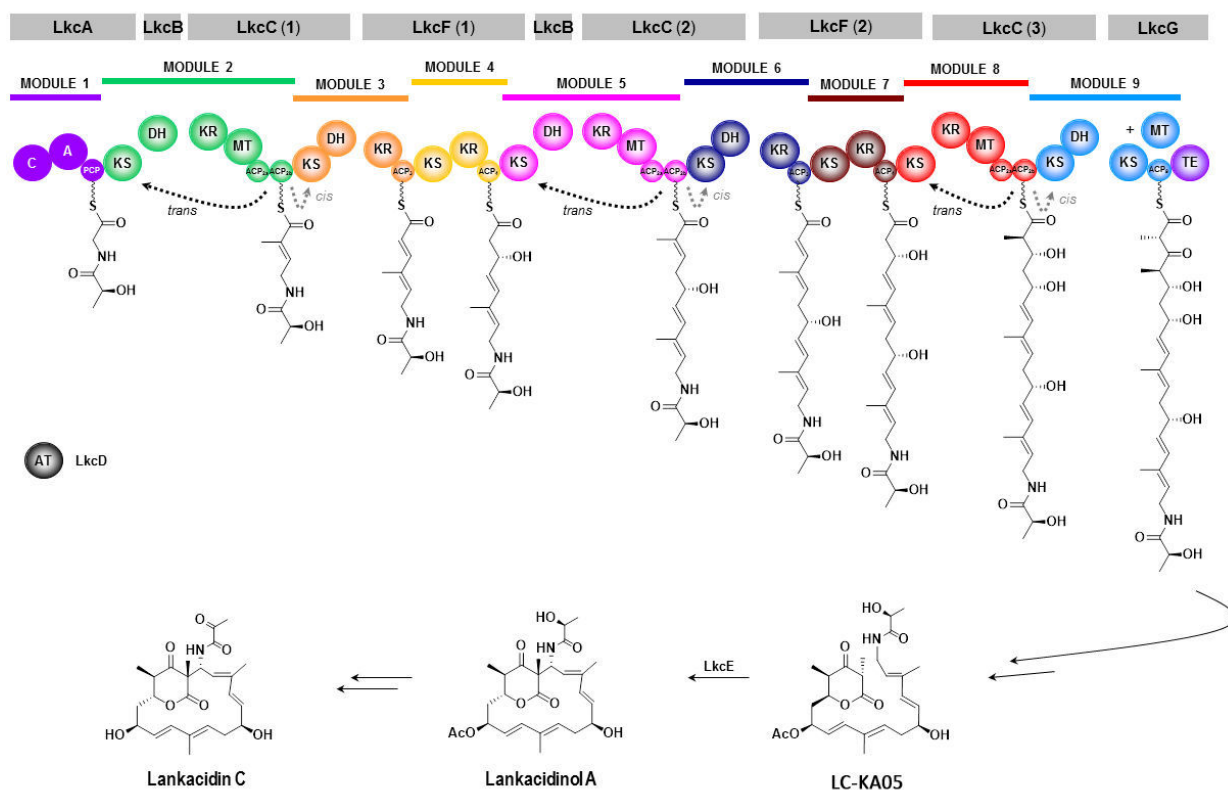


Figure 50 : Schéma représentatif de la voie de synthèse proposée pour la lankacycline C (réalisée par KJ Weissman).

Dans ce modèle, la LkcC et la LkcF interviennent de façon itérative, respectivement trois et deux fois. La libération de la chaîne par le domaine TE du module 9 aboutit à la formation d'un intermédiaire qui est oxydé par une amine oxydase putative, la LkcE, permettant la cyclisation de la molécule. LkcB est une DH discrète qui intervient à deux reprises en agissant en *trans*. L'activité acyltransférase est portée par LkcD, qui interagit en *trans* avec chacun des modules PKS.

La libération de la chaîne par le domaine TE de LkcG n'entraîne pas la cyclisation complète du polykétide. L'intervention d'une enzyme post-PKS est donc nécessaire pour catalyser la cyclisation du composé libéré. Ainsi, la LkcE, une flavoenzyme appartenant à la famille des amine oxydases, catalyse l'oxydation de l'intermédiaire amide non-cyclique en imide, permettant ainsi la cyclisation de la lankacidine par une réaction de type Mannich (Figure 51). Une délétion du gène *lkcE* chez *S. rochei* génère une souche produisant un intermédiaire non cyclisé, LC-KA05, dont la masse est supérieure de deux unités par rapport à celle du produit final, signifiant que deux hydrogènes supplémentaires sont présents sur l'intermédiaire (Arakawa, *et al.*, 2005).

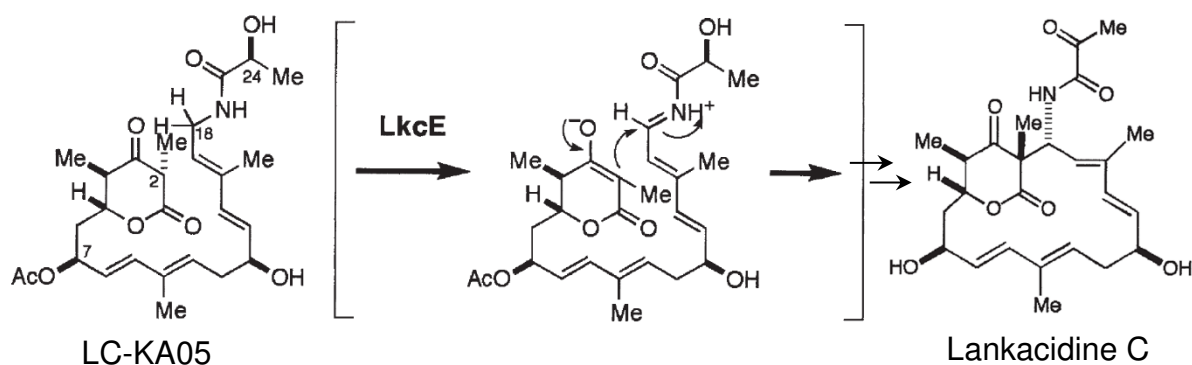


Figure 51 : Mécanisme proposé pour la macrocyclisation de la lankacidine C (adapté de Arakawa, et al., 2005).

La LkcE catalyserait l'oxydation de la fonction amine du LC-KA05 en imine par transfert d'un électron et d'un proton sur le FAD, et utiliserait une catalyse base générale pour subtiliser un proton au niveau du carbone 2. Ce carbone ainsi activé peut réaliser une attaque nucléophile au niveau de l'imine, permettant la cyclisation du polykétide pour former le lankacidinol A. Ce dernier subira ensuite une oxydation à C24 et une élimination du groupement acétate pour former la lankacidine C.

Sur la base d'études portant sur les monoamines oxydases humaines, deux mécanismes ont été proposés pour le transfert d'électron au FAD permettant l'oxydation de l'amine (Fitzpatrick, 2010). Dans le premier, un électron de l'azote de la fonction amine serait transféré au niveau de l'azote réactif du noyau isoalloxazine du FAD passant ainsi de la forme oxydée à la forme semi-quinone. Ensuite, un transfert de proton du carbone de l'amine jusqu'au FAD permettrait la formation de l'imine et du FAD réduit (Figure 52).

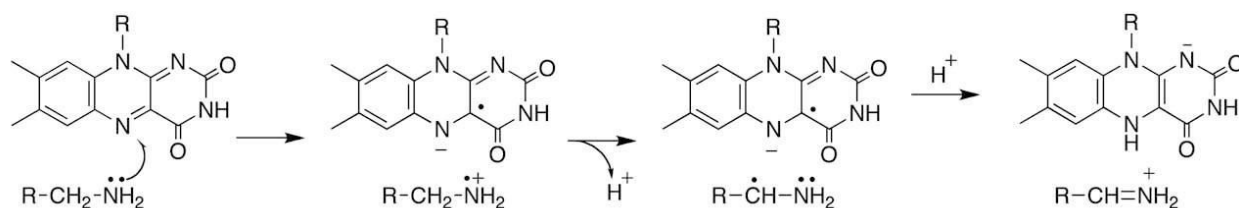


Figure 52 : Mécanisme proposé pour l'oxydation d'une amine par une flavoenzyme (tiré de Fitzpatrick, 2010).

Un autre mécanisme (Fitzpatrick, 2010) impliquerait l'attaque de l'amine sur le noyau isoalloxazine, ce qui aboutirait à la formation d'un intermédiaire covalent. Celui-ci, par un réarrangement moléculaire conduirait à la formation de l'imine et du FAD réduit (Figure 53).

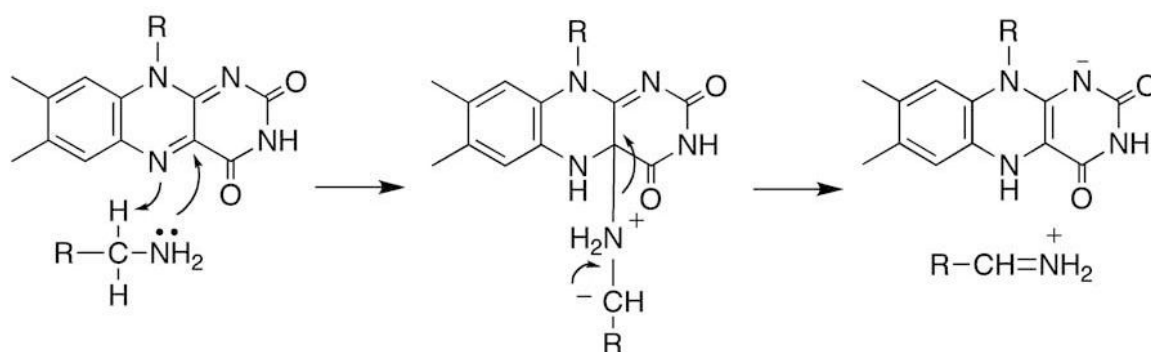


Figure 53 : Deuxième mécanisme proposé pour l'oxydation d'une amine avec intermédiaire covalent entre l'amine et le cycle isoalloxazine du FAD (tiré de Fitzpatrick, 2010).

Un mécanisme alternatif implique un transfert de l'hydrogène en tant qu'hydrure directement sur l'azote du FAD. Cela implique l'absence de catalyse base lors de l'oxydation de l'amine, de même que l'absence d'un intermédiaire carbanion. Cette hypothèse est soutenue par les structures cristallographiques des amines oxydases puisqu'aucun résidu susceptible d'agir comme catalyseur basique n'est présent au sein des sites actifs.

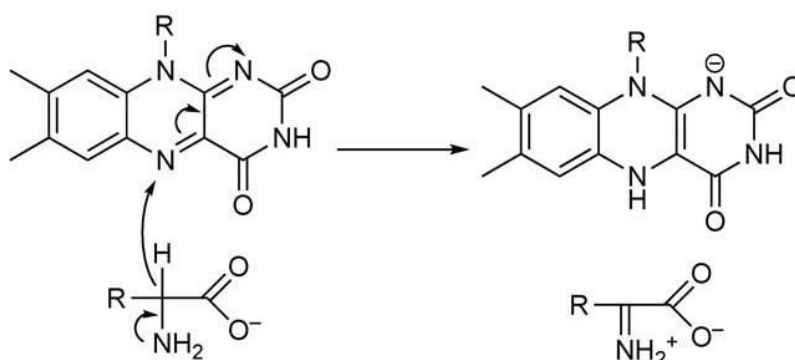


Figure 54 : Mécanisme alternatif proposé pour l'oxydation d'une amine avec transfert d'hydrure depuis l'amine directement sur le FAD (tiré de Fagan, et Palfey, 2010).

La demi-réaction de réoxydation du FAD peut impliquer différents partenaires. Le NAD ou NADP peut intervenir pour réoxyder le co-facteur flavinique, de même que l'oxygène moléculaire. Dans le cas des amines-oxydases, la réoxydation se fait grâce à l'oxygène moléculaire O_2 (Fitzpatrick, 2010). L' O_2 va tout d'abord céder un électron au FAD réduit, permettant ainsi la formation de l'anion superoxyde $O_2^{\cdot-}$ et de la forme semi-quinone du FAD. Les deux espèces radicalaires mettent en commun un électron afin de former une liaison covalente, catalysée par un acide général. Un

réarrangement moléculaire s'opère alors, aboutissant à la formation de peroxyde d'hydrogène et de la forme oxydé du cycle isoalloxazine (Figure 55) (Walsh, et Wencewicz, 2012).

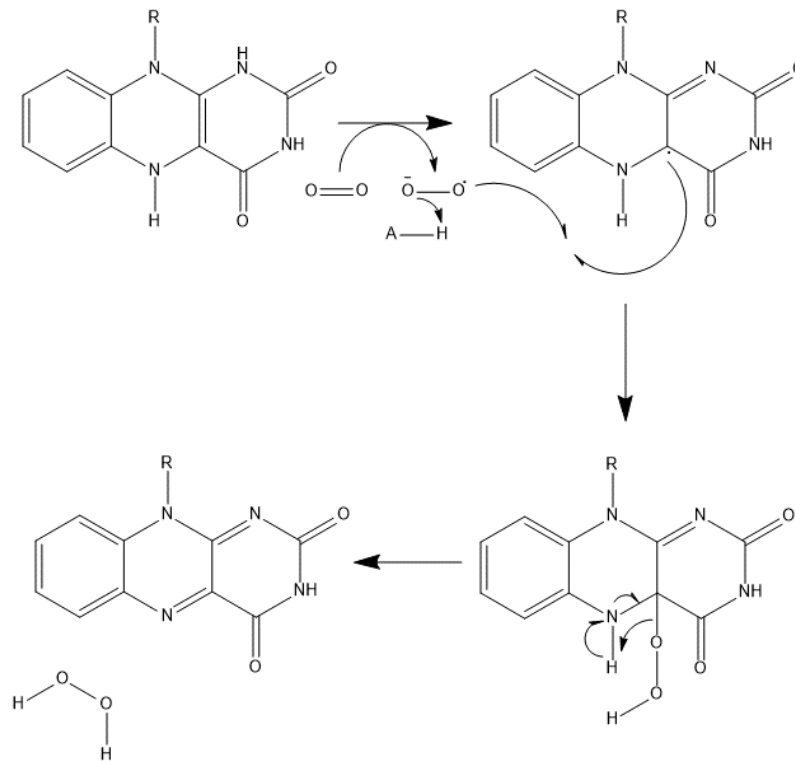


Figure 55 : Mécanisme de réoxydation du FAD par l'oxygène moléculaire O_2 .

OBJECTIFS

Afin que l'ingénierie des PKS connaisse davantage de réussite, il est nécessaire d'obtenir plus de connaissance sur ces systèmes complexes. Appréhender les interactions au sein d'un module PKS se révèle être capitale pour en comprendre le fonctionnement global. Jusqu'alors, seules des structures de domaines ou di-domaines PKS avaient été résolus à haute résolution. Bien que la structure de la mFAS constitue un bon modèle, son mode de fonctionnement itératif induit des divergences importantes par rapport aux PKS modulaires. Ainsi, nous ne disposons que de très peu d'informations sur l'organisation structurale des PKS ou même des modules PKS. Ce manque de connaissance est encore plus criant pour les PKS *trans*-AT, qui constituent pourtant des modèles intéressants pour la biologie synthétique.

Notre but a donc été d'étudier la structure en solution d'un module modèle de PKS *trans*-AT. Pour cela nous avons fait le choix de l'étude du module 5 de la PKS synthétisant la virginiamycine M, un antibiotique dont la fixation au sein du ribosome est bien caractérisée et qui constitue de ce fait un candidat potentiel pour l'approche de biologie synthétique. Ce module consiste en la succession d'un domaine KS et de deux domaines ACP (ACP_{5a} et ACP_{5b}). L'objectif était d'utiliser une approche combinant SAXS, RMN et modélisation par homologie de séquence pour à la fois résoudre la structure du module à faible résolution, mais aussi de disposer des structures à haute résolution ou de modèle à l'échelle atomique des différents domaines afin de les positionner au sein de l'enveloppe SAXS. L'étude de la dynamique du module entier était également un objectif prioritaire afin d'identifier et mieux appréhender les changements de conformations nécessaires à la catalyse.

La compréhension de l'interaction entre sous-unité est aussi une connaissance critique pour la biologie synthétique, puisqu'elle permettrait l'échange de module ou de sous-unité multi-modulaire PKS de façon efficace. Jusqu'alors, l'ensemble des données et structures de DD provenaient de PKS de type *cis*-AT. Lors de cette étude, un domaine situé à l'extrémité C-terminale du module 5 a été identifié. Ce domaine, bien que similaire aux domaines putatifs de dimérisation présent chez certaines PKS de type *cis*, n'entraînait pas la dimérisation du module 5, suggérant qu'il s'agissait également d'un domaine de docking. En effet, un domaine similaire a été identifié à l'extrémité N-terminale du module suivant, le module 6, porté par une sous-unité différente. Cette étude constituerait ainsi la première étude d'une interface d'interaction entre sous-unité d'une PKS de type *trans*-AT. Notre objectif était tout d'abord de mettre

en évidence puis de caractériser l'interaction entre ces deux DD putatifs, et en cas d'interaction, de résoudre la structure par RMN des deux DD en complexe. La finalité était d'identifier les résidus impliqués dans cette interaction, ce qui pourrait permettre leur modification pour la biologie synthétique. Un deuxième objectif était d'étudier la position de l'interface de docking par rapport aux domaines flanquants et donc de voir leur contribution ou non dans l'interaction. Cela permettrait également de replacer cette interface dans le contexte modulaire de la PKS.

Les enzymes post-PKS jouent également un rôle important dans la synthèse du polykétide car ce sont elles qui, en fin de synthèse, lui confèrent son activité biologique. De plus, elles constituent des cibles privilégiées pour la biologie synthétique, puisque ce sont les dernières enzymes à intervenir dans la synthèse du polykétide. Ainsi leur modification entraîne en principe moins de problème de spécificité de substrat. La PKS synthétisant la lankacidine C, un antibiotique de grand intérêt pour la médecine humaine et vétérinaire, produit un intermédiaire linéaire non fonctionnel. Une enzyme post-PKS, la LkcE, décrite comme une amine oxydase potentielle, permet la macrocyclisation du polykétide. Mes objectifs étaient alors de mener à bien une étude structure/fonction de l'enzyme. Pour cela, après avoir mis au point le protocole de production et de purification de l'enzyme, je devais tout d'abord parvenir à cristalliser la protéine native pour démontrer la possibilité d'obtenir des cristaux de l'enzyme, puis sa forme sélénée pour un phasage anomal. Ensuite, je devais co-cristalliser l'enzyme avec deux analogues de substrat mimant chacun une partie du substrat naturel afin d'identifier des résidus catalytiques candidats. Ces résidus seraient ensuite ciblés pour inactiver l'enzyme. Dans le même temps, le Dr. Christophe Jacob procéderait à la délétion du gène codant la LkcE dans *S. rochei* afin d'accumuler le substrat natif de l'enzyme. Une fois extrait, le substrat serait purifié par nos collaborateurs chimistes à Hanovre. Une fois en possession du substrat, je devais parvenir à résoudre la structure de l'enzyme ou d'un de ses mutants en complexe avec son substrat naturel. Je devais aussi mettre au point un test cinétique permettant de déterminer les constantes cinétique de la LkcE. Ainsi nous pourrions proposer un mécanisme catalytique complet pour l'enzyme. L'intérêt de ce sujet est double : disposer d'informations sur des enzymes post-PKS qui peuvent être utilisées en biologie synthétique, mais également caractériser une enzyme capable de catalyser une réaction de cyclisation, très difficile à réaliser en chimie traditionnelle.

*MATERIELS ET
METHODES*

Cette partie matériel et méthodes ne détaillera que les sujets n'ayant pas encore fait l'objet de publication et représente donc un complément aux publications jointes dans ce manuscrit.

I. Matériel biologique

1. Souches bactériennes utilisées

Escherichia coli DH5 α : *supE44, Δ lacU169 (Φ 80 lacZ Δ M15), hsdR17, recA1, endA1, gyrA96, thi-1 et relA1*. Cette souche a été utilisée pour les transformations visant à propager les plasmides, du fait de son taux de transformation élevé.

E. coli BL21 (DE3) : F⁻ *ompT gal dcm lon hsdS_B (r_B⁻ m_B⁻) λ (DE3)*. Cette souche est utilisée pour exprimer certaines protéines recombinantes. Elle dispose dans son génome de la séquence codant l'ARN polymérase du phage T7 sous le contrôle d'un producteur inductible à l'IPTG. Ceci permet l'expression de protéines dont le gène est inséré dans un plasmide de type pET, et placé sous le contrôle du promoteur T7. De plus, cette souche est déficiente pour certaines protéases, ce qui augmente le rendement de la protéine surexprimée.

E. coli Rosetta 2 (DE3) : F⁻ *ompT gal dcm lon hsdS_B (r_B⁻ m_B⁻) λ (DE3) pRARE2 (Cam^R)*. Cette souche, dérivée des BL21, contient le plasmide pRARE permettant la synthèse d'ARN de transfert rares chez *E. coli*.

E. coli B834 pRARE2 (DE3) : F⁻ *ompT hsdS_B (r_B⁻ m_B⁻) gal dcm met λ (DE3) pRARE2 (Cam^R)*. Cette souche, dérivée des Rosetta 2 est auxotrophe pour la méthionine. Elle est utilisée pour la production de protéines sélénées.

E. coli BW25113 : F⁻ *DE(araD-araB)567 lacZ4787(del)::rrnB-3 LAM⁻ rph-1 DE(rhaD-rhaB)568 hsdR514*. Cette souche est utilisée comme hôte dans les expériences de recombinaison entre un cosmide et un produit PCR approprié, en vue du remplacement d'une ORF par une cassette de résistance.

E. coli ET12567 : F⁻ *dam-13::Tn9 dcm6 hsdM hsdR zjj202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44*. Cette souche est utilisée pour la transformation par conjugaison de *S. rochei*.

2. Vecteurs utilisés

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. Ce vecteur permet le clonage de séquences à bouts francs. Il est utilisé afin d'insérer les produits PCR pour séquençage avant le sous-clonage dans des vecteurs d'expression.

pBG102 (Center for Structural Biology, Vanderbilt University) : Ce vecteur d'expression, dérivé du pET27 (Novagen), permet de réaliser une fusion traductionnelle permettant d'obtenir une protéine chimérique avec, en N-terminale, une protéine SUMO tagguée His₆. Le plasmide porte un gène de résistance à la kanamycine (Figure 56A).

pBG106 (Center for Structural Biology, Vanderbilt University) : Vecteur d'expression dérivé du pET27 permettant l'ajout d'une étiquette His₁₀ à l'extrémité N-terminale de la protéine. Il contient un gène de résistance à la kanamycine (Figure 56B).

3. Milieux de culture

Chaque milieu a été autoclavé à 120 °C sous une pression de 1 bar.

Milieu *Lysogeny Broth* ou *Luria Bertani* (LB) : Ce milieu riche est composé de tryptone (10 g/L), d'extrait de levure (5 g/L), de NaCl (5 g/L) et le pH est ajusté à 7,5 par ajout de NaOH. Ce milieu est utilisé pour la croissance bactérienne d'*E. coli* lors de pré-cultures, pour l'expression des protéines recombinantes, mais également utilisé en milieu gélosé pour les transformations par ajout de 17 g/L d'agar.

Milieu 2YT : Ce milieu est légèrement plus riche que le LB puisqu'il est composé de 16 g/L de tryptone, 10 g/L d'extrait de levure, 5 g/L de NaCl. Le pH est ajusté à 7 par ajout de NaOH. Ce milieu est principalement utilisé pour faire croître *E. coli* lors des minipréparations d'ADN plasmidique.

Milieu *super optimal broth with catabolite repression* (SOC) : Il s'agit d'un milieu très riche composé de tryptone (20 g/L), extrait de levure (5 g/L), NaCl (0,5 g/L), KCl (2,5 mM), MgCl₂ (10 mM), MgSO₄ (10 mM), glucose (20 mM) et ajusté à pH 7. Ce milieu sert à régénérer les bactéries compétentes après transformation.

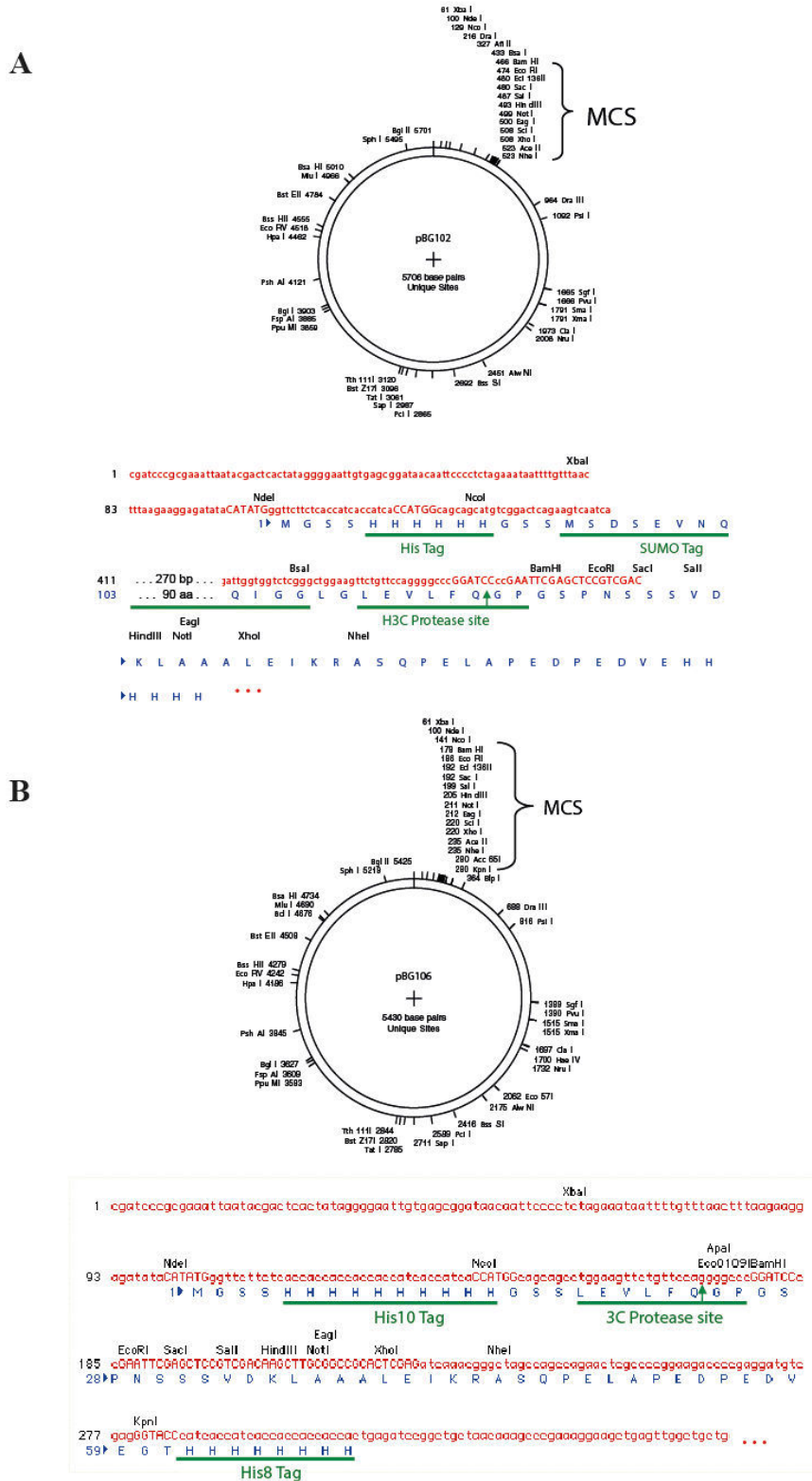


Figure 56 : Représentation schématique des plasmides pBG102 (A) et pBG106 (B).
 Le pBG102 permet d'exprimer des protéines de fusion. Le pBG106 permet de fusionner la séquence d'intérêt avec une étiquette His₁₀ en N-terminale. L'étiquette et la protéine de fusion apportées par ces deux plasmides peuvent être clivées grâce à l'apport d'un site de clivage reconnu par la protéase du rhinovirus humain 3C (h3C). Les sites de restriction *Bam*HI et *Hind*III sont utilisés couramment pour insérer les séquences codant les protéines d'intérêt, excepté pour la *LkcE*, dont la séquence codante a été insérée grâce aux sites *Bam*HI et *Xba*I.

Milieu ZYM 5052 : Ce milieu est préparé avec 10 g/L de tryptone, 5 g/L d'extrait de levure, 25 mM de Na₂HPO₄, 25 mM de KH₂PO₄, 50 mM de NH₄Cl, 5 mM Na₂SO₄, et le pH est ajusté à 7,4. Le milieu est autoclavé et avant la culture 0,5% de glycérol, 0,2% de lactose, 0,05 % de glucose et 1 mM de MgSO₄ sont ajoutés. Il s'agit d'un milieu auto-inductible composé de glucose et de lactose en source carbonée. Les bactéries consomment d'abord le glucose comme source préférentielle de carbone, puis les bactéries vont alors métaboliser le lactose comme source de carbone, ce qui va activer le promoteur lactose et donc induire l'expression du gène codant la protéine d'intérêt.

Milieu M9 : Ce milieu est composé de Na₂HPO₄ (50 mM), KH₂PO₄ (22 mM), NaCl (10 mM) et NH₄Cl (20 mM). Le pH est ajusté à 7,2 par ajout de NaOH. Après autoclave, de la thiamine, de la riboflavine (50 mg/L), du glucose (4 g/L), du CaCl₂ (100 µM) et du MgSO₄ (2 mM) filtrés stérilement sont ajoutés dans le milieu. Il s'agit d'un milieu minimum permettant le contrôle de chaque composant du milieu. Il est utilisé afin d'incorporer des éléments particuliers au sein des protéines synthétisées par les bactéries en culture, et notamment de sélénométhionine lors de la production de la LkcE sélénée (SeLkcE), ou alors de ¹³C et ou ¹⁵N pour les études RMN. Pour la production de la SeLkcE, de la sélénométhionine (40 mg/L), ainsi qu'une solution contenant les 19 autres acides aminés sont ajoutés (40 mg/L).

Les concentrations en kanamycine sont de 50 mg/L de culture dans les pré-cultures et de 5 mg/L de culture dans les cultures permettant l'expression des protéines. Il en va de même pour l'utilisation du chloramphénicol, ajouté à 25 mg/L lors des pré-cultures, et à 2,5 mg/L dans les cultures permettant l'expression de protéines recombinantes. En effet, ces antibiotiques étant des inhibiteurs de la synthèse protéique, la diminution de leur concentration dans les cultures permet d'augmenter le rendement des productions de protéines recombinantes.

4. Solutions d'usage courant

- **Bleu de dépôt pour gel d'agarose (6X)** : Tris HCl 10 mM pH 8 ; EDTA 1 mM ; glycérol 25% (v/v) ; bleu de bromophénol 0,12% (p/v).
- **Bleu de dépôt pour gel de protéines dénaturant** : SDS 3% (v/v) ; glycérol 10% (v/v) ; bleu de bromophénol 0,1 mg/ml ; DTT 100 mM ; Tris-HCl 80 mM, pH 6,8.
- **Marqueur de taille de protéines MiMoPA** : T7 ARN polymérase (97 kDa), albumine de sérum bovin (66 kDa), rétinaldéhyde déshydrogénase (53 kDa), ovalbumine glycosylée (44,8 kDa), ovalbumine (42,8 kDa), glycéraldéhyde-phosphate déshydrogénase (37,8 kDa), MsrA (24 kDa), thiorédoxine (12 kDa).
- Solution I de coloration des gels SDS-PAGE : Ethanol 50% ; acide acétique 10%.
- Solution II de coloration des gels SDS-PAGE : Ethanol 5% ; acide acétique 7,5%.
- Solution stock de coloration des gels SDS-PAGE : Bleu de Coomassie G250 0,25% ; éthanol 96%.
- **Tampon de migration SDS-PAGE** : glycine 14,25 g/l ; Tris 3 g/l ; SDS 1 g/l (v/v).
- **Tampon TAE de migration en gel d'agarose** : Tris 40 mM, acétate de sodium 5 mM, pH 8,0 ; EDTA 1 mM.

II. **Sous-clonage, production et purification des protéines recombinantes**

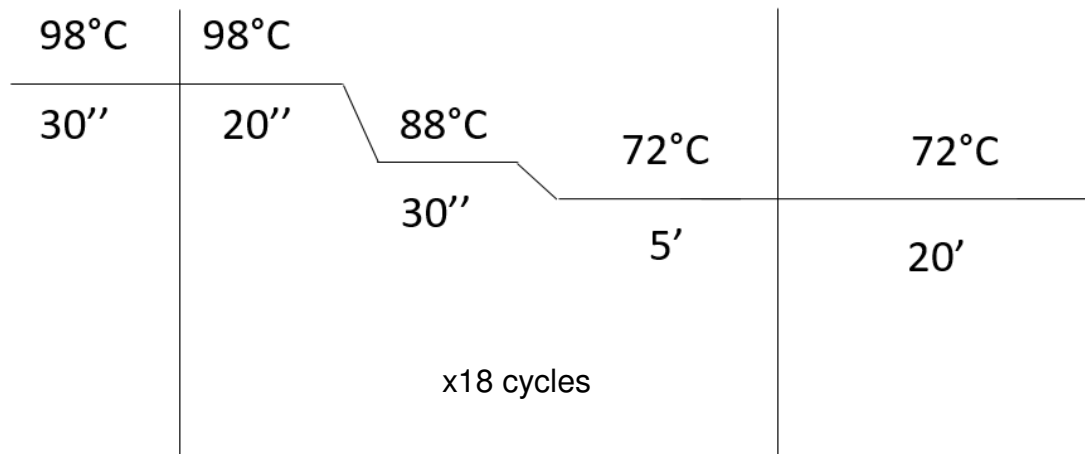
1. Sous-clonage de la séquence codant la LkcE

L'équipe du Pr Peter Leadlay (Université de Cambridge) nous a fourni une souche d'*E. coli* transformée à l'aide du plasmide pGEX-LcsJ contenant la séquence codante de la LkcE flanquée de sites de coupure par *Bam*HI et *Xho*I (A noter : LcsJ est la nomenclature utilisée au laboratoire du Pr Leadlay pour la LkcE). Après culture en milieu 2YT, l'ADN plasmidique a été purifié par minipréparation en utilisant le kit Nucleospin® plasmid (Macherey Nagel). Le plasmide a été ensuite digéré par *Bam*HI et *Xho*I, le fragment correspondant à la séquence d'intérêt a été purifié sur gel d'agarose, puis cloné dans le vecteur pBG102 préalablement digéré par les mêmes enzymes. Le plasmide recombinant a été amplifié dans la souche DH5α d'*E. coli*.

Après minipréparation plasmidique, la séquence codant la LkcE a été vérifiée par séquençage (GATC Biotech).

2. Mutagenèse dirigée

Dans le but de co-cristalliser la protéine avec son substrat naturel, il est nécessaire d'empêcher la catalyse. Des acides aminés au site actif de la LkcE identifiés comme potentiellement catalytiques ont alors été mutés. Les mutants ont été obtenus par mutagenèse dirigée en « *rolling-circle* », en utilisant des oligonucléotides complémentaires. La matrice PCR utilisée était le pBG102-LkcE. Suite à la PCR (Figure 57), 1 μ L de *DpnI* Fast digest (Thermo Fisher Scientific) a été ajouté à 30 μ L de produit PCR dans les conditions recommandées par le fournisseur. La réaction a alors été placée 30 minutes à 37°C. Cette étape a permis d'éliminer l'ADN parental et de ne conserver que l'ADN muté pour la transformation bactérienne. Après transformation, les cellules DH5 α d'*E. coli* ont été étalées sur milieu sélectif. Les plasmides ont été extraits par minipréparation d'ADN plasmidique et ensuite analysés par séquençage.



Oligonucléotides :

- Y182F sens : GACTACAACCTCATGITCGTCGCGCTGTCGTT
- Y182F anti-sens : AACGACAGCGCGACGAACATGAGGTTGTAGTC
- R326Q sens : ATGGGCAGCCTGACCCAGAACTGCAACGTGCTC
- R326Q anti-sens : GAGCACGTTGCAGTTCTGGGTCAGGCTGCCCAT
- R326L sens : ATGGGCAGCCTGACCCTCAACTGCAACGTGCTC
- R326L anti-sens : GAGCACGTTGCAGTTGAGGGTCAGGCTGCCCAT

ADN polymérase utilisée : Phusion (Finnzymes Ozyme)

Additif utilisé : 10% Diméthyl sulfoxyde (DMSO)

Figure 57 : Protocole PCR utilisé pour la mutagenèse dirigée de la LkcE.
Les nucléotides soulignés correspondent aux codons qui ont été mutés.

3. Production et purification de la LkcE et de ses mutants

A) Culture permettant l'expression de la LkcE

La LkcE a été produite dans des cellules Rosetta 2 d'*E. coli* transformées par le plasmide pBG102 contenant la séquence codant l'enzyme. Elle a donc été exprimée sous forme de protéine de fusion avec une protéine SUMO tagguée His₆ en C-terminal. Après une pré-culture en milieu LB, d'une nuit à 37 °C, la culture des bactéries exprimant cette protéine est faite en milieu LB supplémenté en riboflavine (1 g/L) afin de favoriser la production de FAD. Pour induire l'expression de protéines chaperonnes

qui aident au repliement de la protéine, mais également pour ralentir le temps de génération des bactéries et ainsi favoriser la production et l'incorporation du FAD dans la protéine, les cultures ont été placées à 37 °C jusqu'à l'obtention d'une DO_{600nm} de 0,6. Un choc chimique et thermique a été alors réalisé en ajoutant 3% d'éthanol à la culture qui a été alors placée à 4°C durant 2 h. La production de la LkcE a été alors induite par ajout d'IPTG (0,1 mM) et la culture a été placée à 15 °C sur la nuit.

La culture a été ensuite centrifugée durant 30 min à 3500 *g*. Les bactéries ont été ensuite resuspendues dans 30 mL par litre de culture de tampon phosphate-NaCl (NaPi) 100 mM, glycérol 10%, EDTA 10 mM, pH 7,4. Elles ont été alors centrifugées à nouveau, le surnageant a été éliminé, et les culots bactériens ont été congelés à -20 °C.

B) Lyse cellulaire et purification de la LkcE

a. Lyse cellulaire

Les culots cellulaires de chaque litre de culture ont été repris dans du tampon HEPES 30 mM, NaCl 500 mM, pH 7,5 à raison de 10 mL par unité de DO_{600nm} en fin de culture. Quatre cents unités de benzonase pour 100 mL de surnageant de lyse cellulaire ont été ajoutées, ainsi que 6 mM de $MgSO_4$ pour éliminer les acides nucléiques. Les cellules ont été lysées au désintégrateur de cellule de type French Press (Basic Z, Constant Systems Ltd.) à 15 kPsi (1000 bars) à 4 °C. Trente mL de tampon de lyse ont été ajoutés dans la French Press afin de s'assurer de ne pas perdre d'échantillon. Les débris cellulaires ont été éliminés par centrifugation à 17 000 *g* pendant 40 min

b. Chromatographie d'affinité Nickel-Sépharose

Après filtration sur filtre de 0,22 μm , et addition de 70 mM d'imidazole, le surnageant de lyse a été injecté à 3 mL/min sur colonne Ni-Sépharose (GE Healthcare Life Sciences) couplée à un système Äkta Avant (GE Healthcare Life Sciences). La colonne a été préalablement équilibrée avec du tampon HEPES 30 mM, NaCl 500 mM, imidazole 70 mM, pH 7,5. La protéine présente une affinité pour les groupements nickel en raison de l'étiquette His₆ en N-terminal de la fusion SUMO apportée par le

plasmide pBG102. La protéine a été ensuite éluée par compétition en utilisant un palier à 250 mM d'imidazole, à un débit de 5 mL/min. Les fractions collectées ont été analysées par SDS-PAGE 12,5% révélé au bleu de Coomassie et les fractions contenant la LkcE ont été regroupées.

c. Protéolyse de la protéine de fusion et du tag histidine suivie d'une chromatographie d'affinité Nickel-Sépharose inverse

Les différentes fractions rassemblées ont été mises en présence de 150 µg de protéase du rhinovirus humain h3C et mises à dialyser sur la nuit contre 5 L de tampon de lyse pour éliminer l'imidazole.

Après clivage de la protéine SUMO tagguée His₆, l'échantillon a été injecté à nouveau sur colonne Nickel-Sépharose (GE Healthcare Life Sciences) à un débit de 2 mL/min. La colonne a été préalablement équilibrée avec du tampon HEPES 30 mM, NaCl 500 mM, pH 7,5 contenant 20 mM d'imidazole. La LkcE a donc été éluée dans la fraction non retenue. Après la nickel inverse, la protéine a été diluée 3 fois dans du tampon HEPES 30 mM, EDTA 1 mM, pH 7,5 afin de diminuer la force ionique du tampon.

d. Chromatographie échangeuse d'anions

Une colonne Q-sépharose (agarose 6% sur lequel sont greffés des groupements triméthylammonium chargés positivement) couplée à un système Äkta Avant (GE Healthcare Life Sciences) a été équilibrée avec du tampon HEPES 30 mM, NaCl 100 mM, EDTA 1 mM, pH 7,5. L'échantillon dilué a été injecté à un débit de 5 mL/min. Un gradient croissant de NaCl (de 100 mM à 1 M) a été utilisé avec un débit de 5 mL/min pour éluer la protéine. Les fractions collectées ont été analysées par gel SDS-PAGE 12,5%, et les fractions contenant la LkcE ont été regroupées.

e. Chromatographie d'exclusion stérique

Afin d'éliminer les éventuels derniers contaminants et les agrégats de protéines, une étape de filtration sur gel a été effectuée. Pour cela, la protéine a été tout d'abord concentrée à l'aide d'un concentrateur Amicon Ultracel-30 (Merck Millipore) par centrifugation à 4000 g jusqu'à un volume inférieur à 8 mL. La protéine a été ensuite injectée à l'aide d'une boucle d'injection de 10 mL sur une colonne Superdex 200 26/60

prep grade (GE Healthcare Life Sciences) couplée à un système Äkta FPLC (GE Healthcare Life Sciences). La colonne a été préalablement équilibrée avec du tampon HEPES 30 mM, NaCl 150 mM, EDTA 1 mM, pH 7,5. L'élution a été réalisée avec un volume colonne de ce même tampon à un débit de 2 mL/min. Les protéines ont été séparées suivant leur diamètre hydrodynamique et les différentes fractions de 2 mL ont été analysées par gel SDS-PAGE 12,5 % révélé au bleu de Coomassie. Les fractions contenant la protéine pure ont été regroupées. La protéine a été concentrée jusque 20 mg/mL par centrifugation à 4000 *g* avec un concentrateur Amicon Ultracel-30 (Merck Millipore) et congelée en fractions aliquotes de 50 μ L dans l'azote liquide. Elle est ensuite stockée à -80 °C.

Un dosage spectrophotométrique en UV-visible permet d'estimer le pourcentage de protéine ayant incorporé le co-facteur flavinique. Pour cela l'absorbance de la protéine (280 nm, $\epsilon = 49\ 333,5$) et l'absorbance du FAD (450 nm, $\epsilon = 11\ 300$ nm) ont été mesurées à l'aide d'un spectrophotomètre SAFAS (UVmc²).

4. Production de la SeLkcE

Afin de produire la LkcE sous forme sélénée (SeLkcE) en forçant l'incorporation des sélénométhionines, nous avons utilisé un protocole de double répression de la production de méthionine. Pour cela, la souche B834 pRARE2 d'*E. coli* a été transformée avec le plasmide pBG102 comportant la séquence codant la LkcE. Les bactéries ont ensuite été mises en culture en milieu M9 dans lequel a été ajouté un mélange composé des 19 acides aminés naturels (sauf la méthionine) et de sélénométhionine (Hendrickson, *et al.*, 1990). Le reste du protocole de production et purification est identique à celui de la LkcE native et de ses mutants.

III. Production, extraction et purification du substrat de la LkcE.

Afin d'isoler le substrat de la LkcE, il a fallu tout d'abord inactiver l'enzyme qui le métabolise. Le Dr. C Jacob a réalisé une délétion du gène codant la LkcE dans la souche productrice de la lankacidine C : *S. rochei*. Pour ce faire, un couple d'oligonucléotides a été conçu pour amplifier la cassette de résistance à l'apramycine,

avec en 5', 39 nucléotides situés en amont du gène codant la LkcE, et en 3', 39 nucléotides situés en aval du gène codant la LkcE. Une PCR en utilisant comme matrice le plasmide pIJ773 (Figure 58A) a permis d'amplifier la cassette contenant le gène de résistance à l'apramycine tout en incorporant les séquences flanquantes du gène codant la LkcE (Figure 58B). Ce produit PCR a ensuite été introduit dans le cosmide Lc2B12 fourni par le Pr. P.F. Leadlay (Université de Cambridge), dérivé du SuperCos1 (Figure 58C) (Evans, *et al.*, 1989 ; Dickschat, *et al.*, 2011). Pour cela, des cellules BW25113 d'*E. coli* ont été co-transformées par le cosmide Lc2B12, le plasmide pIJ790 (Figure 58D) et le fragment PCR. Le plasmide pIJ790 permet l'expression de la recombinase λ Red qui rend possible une recombinaison homologue entre les extrémités du fragment PCR et le cosmide Lc2B12 au niveau de la séquence codant la LkcE. Les clones pour lesquels une recombinaison homologue a eu lieu ont été sélectionnés sur milieu supplémenté en apramycine. Le cosmide ayant incorporé la cassette de résistance à l'apramycine a été alors extraite par minipréparation d'ADN plasmidique puis envoyé en séquençage.

Des bactéries ET12567 d'*E. coli* ont été ensuite co-transformées par le cosmide et le plasmide pUZ8002 (Figure 58E) aidant à la conjugaison bactérienne. Ces bactéries ont servi pour la conjugaison avec la souche *S. rochei* au stade de spores. Il y a alors eu recombinaison homologue entre le cosmide et le plasmide PSLA2-L de *S. rochei* qui code naturellement le cluster de gènes de la PKS synthétisant la lankacidine C. Ceci entraîne par conséquent le remplacement du gène codant la LkcE sur le plasmide PSLA2-L par la cassette de résistance à l'apramycine.

La souche ainsi a été mise en culture dans le milieu YMG (4 g de glucose, 4 g d'extrait de levure, 10 g d'extrait de malt par litre de culture, pH 7,2 ajusté avec du KOH) pendant environ 72 h à 30 °C sous agitation à 200 rpm. Trois cultures de 10 litres chacune ont été nécessaires à l'obtention du substrat. Après centrifugation à 4000 rpm durant 30 min, le milieu de culture a été extrait à l'acétate d'éthyle (3 fois 1 équivalent volume). La phase organique a été conservée et évaporée sous vide à l'aide d'un système rotavapeur (40 °C sous vide).

A partir de 10 L de milieu de culture, entre 100 mg à 200 mg d'extrait total ont été récupéré. Le produit de l'extraction a ensuite été envoyé au laboratoire du Pr. A Kirschning à Hanovre afin de purifier le substrat.

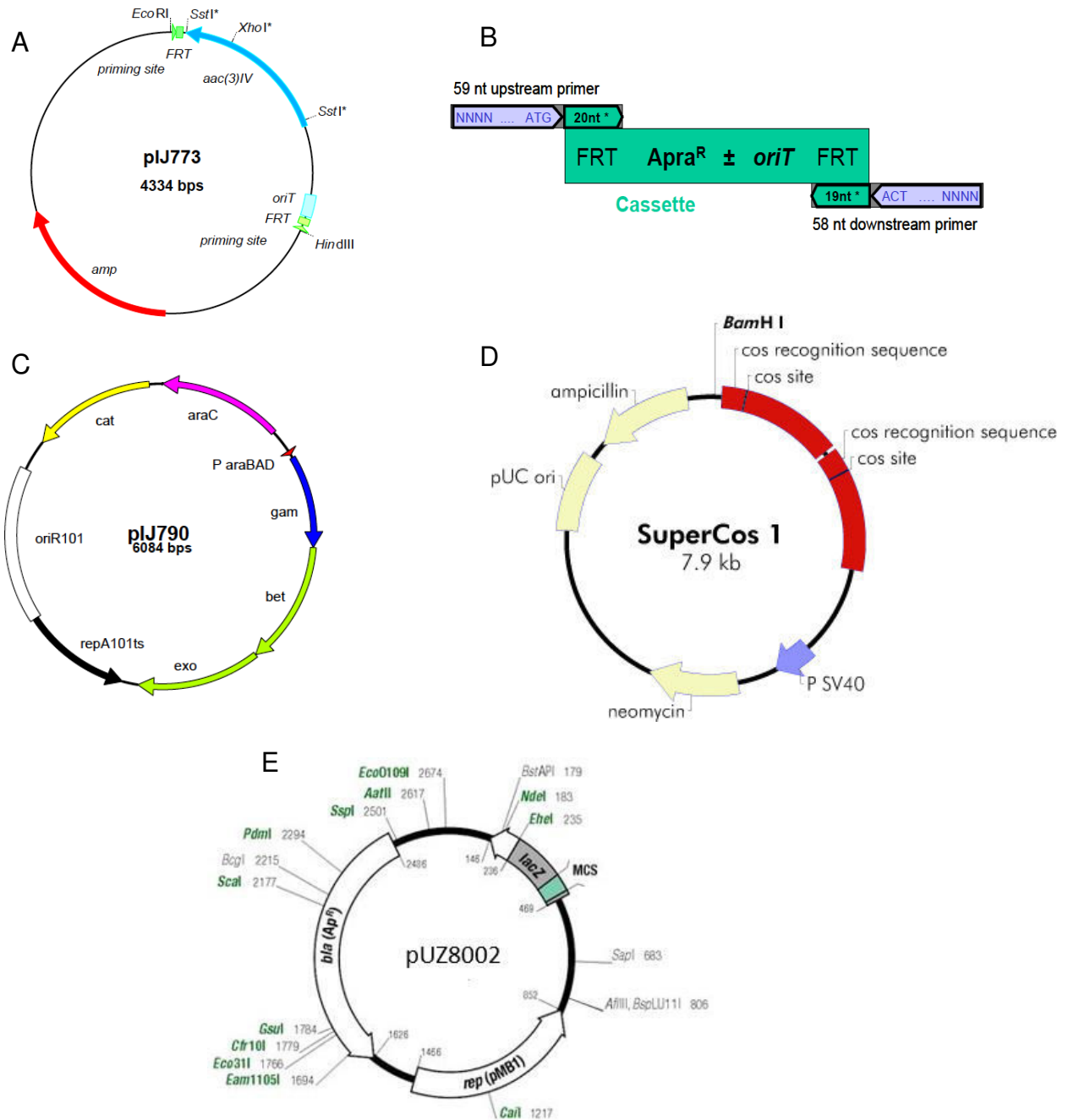


Figure 58 : Constructions utilisées pour déléter le gène codant la LkcE chez *S. rochei*.
A : Plasmide pIJ773 ayant servi comme matrice PCR pour l'amplification de la cassette de résistance à l'apramycine.
B : Fragment amplifié par PCR : Les séquences en bleu correspondent aux 39 et 38 nucléotides apportés par les oligonucléotides, correspondant respectivement à la séquence en amont et en aval du gène codant la LkcE. FRT (flippase recognition target) est une séquence permettant l'excision de la cassette de résistance, si besoin.
C : Plasmide pIJ790 favorisant la recombinaison homologue chez *E. coli*.
D : Cosmide SuperCos1 ayant servi de base à l'obtention du cosmide Lc2B12 par l'équipe du Pr. P.F. Leadlay.
E : Plasmide pUZ8002 favorisant la conjugaison bactérienne entre *E. coli* et *S. rochei*

IV. Etudes préliminaires de la LkcE

1. Homogénéité en solution

La diffusion dynamique de la lumière (DLS) permet de déterminer l'état d'agrégation de la protéine ainsi que la taille des particules présentes en solution. La mesure de la variation d'intensité de diffusion des particules en solution dépend de la diffusion des solutés en solution qui est fonction du rayon hydrodynamique. Avant la mesure, l'échantillon a été centrifugé à 15000 *g* pendant 30 min pour éliminer les éventuelles poussières. Les mesures ont été effectuées sur un appareil Zetasizer NanoS (Malverne) et chaque mesure comportait une dizaine d'acquisitions de 1 s à 25 °C.

2. Vérification de l'intégrité de la LkcE et caractérisation du co-facteur flavinique par spectrométrie de masse

L'échantillon de LkcE purifié a été dessalé par chromatographie liquide (HPLC, Äkta Explorer, GE Healthcare) sur une colonne en phase inverse C8 (Grace) et élué en gradient croissant d'acétonitrile, contenant du TFA 0,1%. Le pic correspondant à la protéine et celui correspondant à la flavine ont ensuite été analysés par spectrométrie de masse (F. Dupire, Service Commun de Spectrométrie de masse de la Faculté des Sciences et Technologie de l'Université de Lorraine).

V. Etude de la LkcE par cristallographie aux rayons X

La cristallographie aux rayons X est une technique de référence pour la détermination de structures tridimensionnelles de protéines à l'échelle atomique. Cette technique se divise en trois phases : 1) la cristallisation de la protéine. En effet, l'étude cristallographique nécessite d'obtenir des cristaux de protéine d'une taille et d'une qualité suffisante pour obtenir une diffraction ; 2) l'enregistrement et le traitement de données de diffraction des rayons X ; et 3) la résolution de la structure à partir des données de diffraction, qui nécessite une étape de phasage et un processus itératif d'affinement de la structure.

1. La cristallisation de la LkCE

A) Principes et techniques de cristallisation

La cristallisation consiste à faire passer la protéine d'un état libre en solution à un état cristallin qui est un état solide et ordonné périodiquement. En effet dans un cristal, les molécules sont orientées et disposées à l'identique d'une maille à l'autre. Pour aboutir à un état cristallin, la solubilité de la protéine est diminuée lentement grâce à des agents précipitant afin de favoriser les interactions spécifiques intermoléculaires (Figure 59).

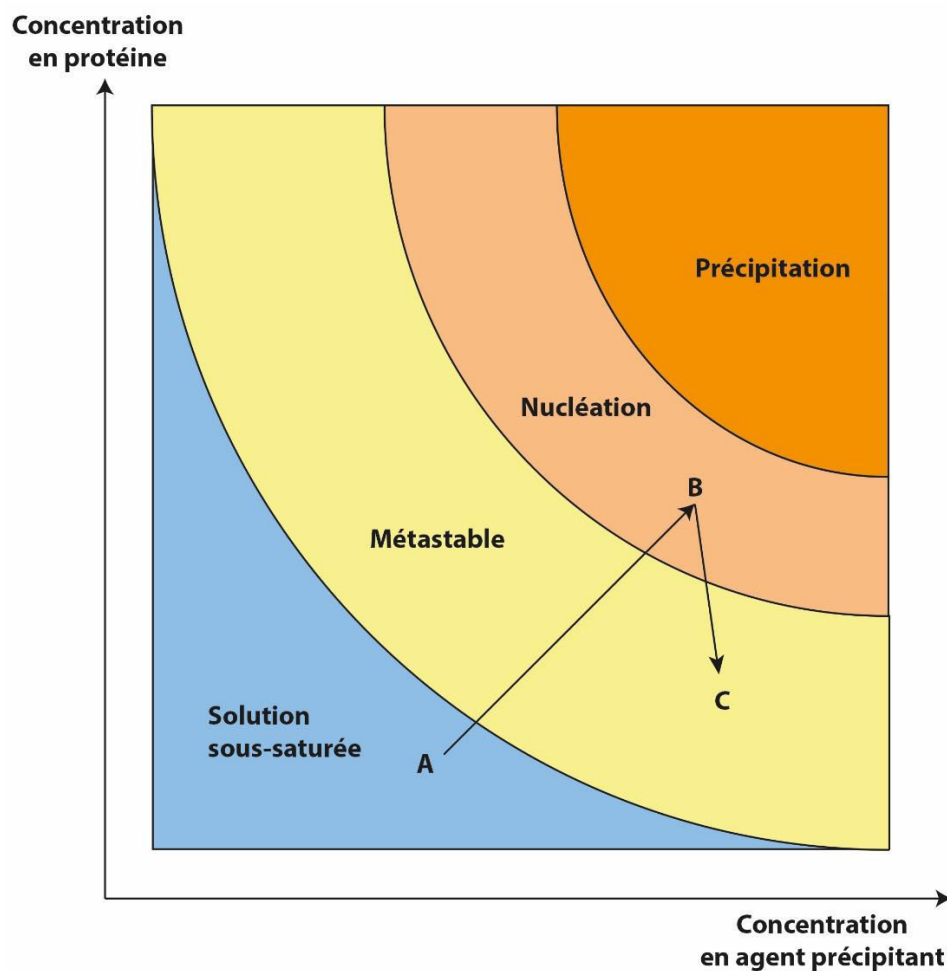


Figure 59 : Diagramme de phase d'une protéine.

Afin de faire passer la protéine de l'état soluble à un état cristallin, on augmente peu à peu sa concentration, de même que la concentration en agent précipitant. Ainsi la protéine, qui est en solution sous-saturée (A), va passer progressivement vers une zone de nucléation (B). C'est à cette concentration que vont se former les premiers germes cristallins. La protéine passant petit à petit d'un état libre en solution à un état cristallin, la concentration de celle-ci va diminuer dans la solution. Ainsi la croissance du cristal va se poursuivre dans la zone que l'on appelle métastable, qui est une solution sursaturée, qui ne permet pas la formation de germes cristallins mais uniquement leur croissance (C).

Le but va donc être de trouver un compromis entre concentration en protéine et concentration en agent précipitant afin d'amener la protéine dans la zone de nucléation, sans pour autant faire précipiter la protéine. La recherche des conditions de cristallisation est empirique, et s'effectue sous forme d'essais-erreurs en raison des multiples paramètres qui influencent la cristallisation (McPherson, et Gavira, 2014) :

- la pureté de la protéine : dans le cas général, la solution protéique doit être le plus parfaitement homogène possible. Un autre facteur important est le temps de purification de la protéine, pendant lequel la protéine peut se dégrader.
- la concentration en protéine : comme décrit précédemment il s'agit d'un facteur crucial pour la cristallisation de la protéine puisqu'elle régit les interactions protéine-protéine.
- la nature des agents précipitants et leur concentration : ils vont permettre de diminuer la solvataion des protéines et ainsi favoriser les interactions spécifiques.
- le pH : qui va à la fois déterminer la stabilité de la protéine (l'intégrité de la structure) mais aussi l'état de charge de la protéine.
- la présence ou non de ligands : ils peuvent induire un changement conformationnel pouvant rigidifier des parties flexibles de la protéine et donc, aider à la cristallisation.
- l'ajout d'additifs : leur nature peut être très variée (ions, alcools, détergents, peptides, sucres, etc...). Ils peuvent favoriser la cristallisation de la protéine.
- la température : qui affecte la stabilité, la solubilité de la protéine et a une influence majeure sur la cinétique de cristallisation.

La solubilité de la protéine dépend de ses interactions avec le solvant. Le but va donc être de perturber ces interactions protéines-solvant pour désolvater la protéine et forcer les interactions spécifiques protéine-protéine. Pour cela divers agents précipitants sont utilisés :

- les polymères organiques qui agissent par exclusion du solvant au voisinage de la protéine (le polyéthylène glycol (PEG) est le plus utilisé).
- les solvants organiques qui diminuent la constante diélectrique du milieu, favorisant ainsi les interactions ioniques.

- les sels non chaotropiques qui modifient la force ionique. A de faibles concentrations, ils vont favoriser la solubilité de la protéine (*salting in*) mais à des concentrations de l'ordre du molaire, ils vont exclure le solvant et ainsi favoriser les interactions hydrophobes (*salting out*).

Plusieurs techniques permettent de diminuer progressivement la solubilité de la protéine afin de l'amener à un état cristallin. La plus utilisée est la diffusion en phase vapeur, mais cela peut également se faire par dialyse, ou par *batch* (Ducruix, et Giegé, 1999). La technique de diffusion en phase vapeur la plus utilisée est la goutte suspendue (Wlodawer, et Hodgson, 1975). Cette méthode consiste à déposer dans un récipient (appelé couramment puit), un volume (1 mL) de solution contenant l'agent précipitant. Sur une fine lamelle est ensuite déposé un volume (de 1 μL à 5 μL) de solution protéique, auxquels est ajouté un volume de solution précipitante (1 μL à 5 μL). La lamelle est ensuite posée au-dessus du récipient et est scellée hermétiquement. La solution précipitante est donc diluée sur la lamelle par rapport à celle présente dans le puit et donc une diffusion de vapeur va s'établir de la goutte vers le puit afin d'équilibrer les concentrations. Les concentrations en agent précipitant et en protéine vont donc progressivement augmenter et, dans des conditions favorables, la protéine va atteindre la zone de nucléation et ainsi former les premiers germes cristallins (Figure 60). Cette technique a été utilisée en boîte *Linbro* pour la reproduction des cristaux et l'affinement des conditions de cristallisation

La méthode de la goutte assise a également été utilisée pour l'utilisation des *screens* commerciaux permettant de tester de nombreuses conditions de cristallisation. Cette technique de diffusion en phase vapeur diffère de la goutte suspendue par le fait que la goutte de cristallisation est déposée sur un support situé au-dessus ou à côté du puit.

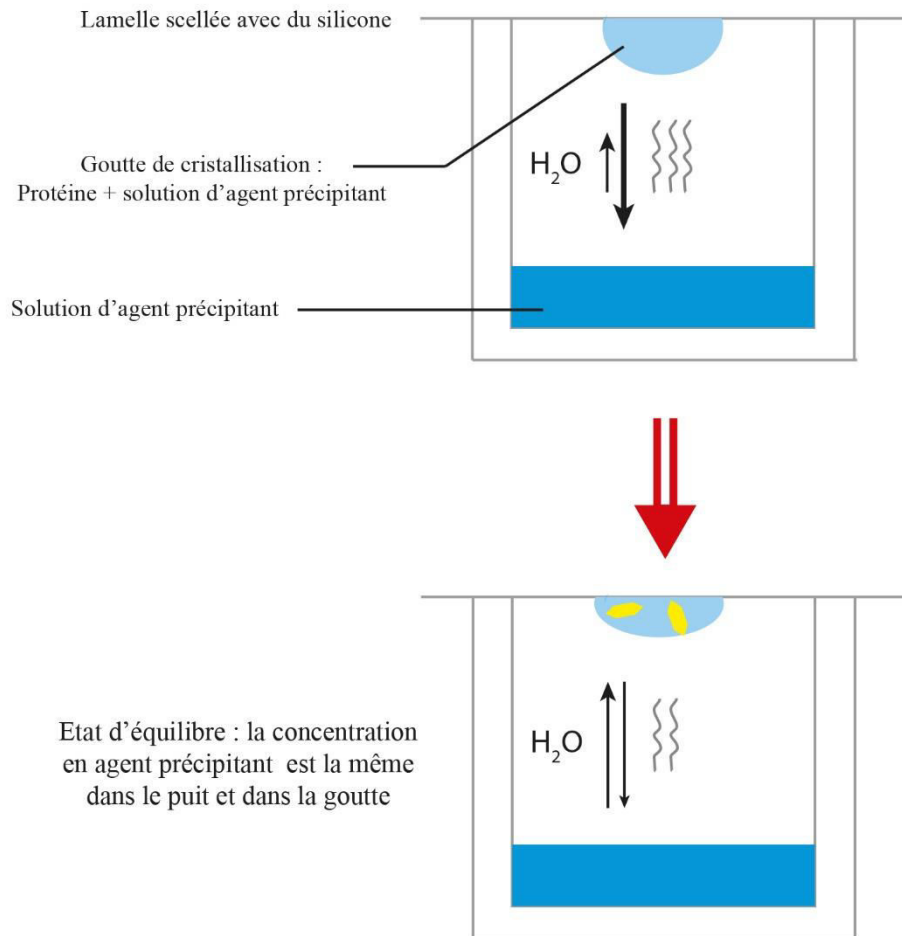


Figure 60 : Principe de cristallisation par diffusion de phase vapeur en goutte suspendue.

Un grand volume de solution d'agent précipitant (1 ml) est placé dans le puit. La goutte de cristallisation est composée d'un certain ratio de solution protéique et de solution d'agent précipitant : trois ratios ont été classiquement utilisés pour chaque condition testée : 1/1 ; 1/2 et 2/1 (en μL). Du fait de la différence de concentration en agent précipitant entre les gouttes et le puit, une diffusion d'eau de la goutte vers le réservoir va se produire. Ainsi la concentration en protéine et en agent précipitant dans la goutte va progressivement augmenter jusqu'à atteindre un état d'équilibre.

B) Recherche et amélioration des conditions de cristallisation de la LkCE

a. Criblage de conditions de cristallisation

Les premiers tests de cristallisation ont été réalisés avec des *screens* commerciaux afin d'obtenir les premiers cristaux et de connaître le comportement de la protéine dans différentes conditions de pH et de force ionique mais aussi l'agent précipitant à utiliser de préférence. Ces *screens* commerciaux sont composés de conditions ayant déjà permis l'obtention de cristaux de protéines et offrent l'avantage d'explorer une grande variété d'agents précipitants (sels, polymères, solvants organiques), de pH, et d'additifs (Jancarik et Kim, 1991).

Ces tests ont été réalisés dans des boîtes de type Greiner MRC 2 (Hampton Research) dans lesquelles chacun des 96 puits est associé à trois cupules permettant la réalisation de trois ratios différents : 1/2, 1/1, et 2/1 (Figure 61). Ces boîtes ont été remplies à l'aide de deux robots différents : un robot Evo 150 (Tecan) pour remplir le puit avec les solutions de cristallisation et un robot Mosquito (TTP Labtech) pour déposer la solution protéique et la solution de cristallisation dans chaque cupule. En effet, le robot Mosquito permet l'utilisation de très faibles volumes (100 et 200 nL).

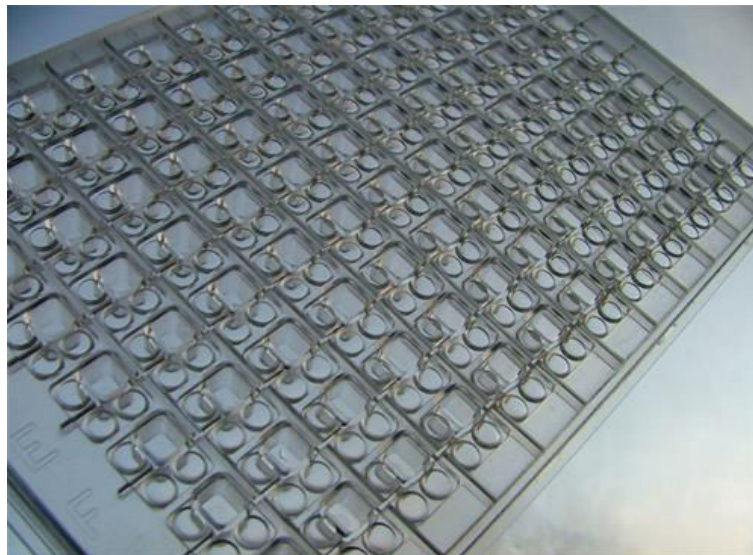


Figure 61 : Boîte Greiner MRC 2 (Hampton Research) utilisé lors des *screens* de conditions de cristallisation.

Les 96 puits sont remplis avec les différentes solutions de cristallisation et les trois cupules permettent de réaliser trois ratios différents.

Pour la LkcE, les *screens* commerciaux JCSG-plus (Molecular Dimensions), Morpheus (Molecular Dimensions), Wizard (Emerald BioSystems), SaltRX (Hampton research), PGA Screen (Molecular Dimensions), PEG/Ion Screen (Hampton research), HR (Hampton research), et Index (Hampton research) ont été utilisés. Pour la SeLkcE seuls les *screens* Index (Hampton research), HR (Hampton research), Wizard (Emerald BioSystems) et PEG/Ion Screen (Hampton research) ont été effectués. Les boîtes ont été ensuite scellées hermétiquement à l'aide d'un film plastique transparent, placées à 20 °C et l'évolution des gouttes a été suivie à l'aide d'une loupe binoculaire.

b. Amélioration des cristaux

Lorsqu'une condition de cristallisation entraîne l'apparition de cristaux, elle a été reproduite dans des volumes plus importants, en goutte suspendue en boîte *Linbro*. Pour cela, des gammes faisant varier pas à pas les concentrations des différents composants (agents précipitants, sels, additifs) de la condition du *screen* commercial ou le pH ont été effectuées, permettant soit l'amélioration des cristaux, soit de comprendre quels sont les composants influençant favorablement la cristallisation.

Une autre méthode consiste à élaborer une série d'essais permettant de tester un maximum de conditions en un minimum d'expériences. Ce sont les plans factoriels incomplets qui combinent entre elles toutes les différentes variables (précipitants, sels, pH, etc...) auxquelles sont associés plusieurs niveaux (les différentes concentrations, valeurs de pH, etc...) (Carter, et Carter, 1979). Cette combinaison s'effectue de manière aléatoire mais de façon à ce que chaque variable soit représentée un nombre identique de fois et en évitant les redondances. Ces plans d'expérience ont été calculés en utilisant le logiciel SAmBA (*Simulated Annealing and BAcktracking methods*) (Audic, et al., 1997).

Lorsque les cristaux sont de tailles insuffisantes, que la nucléation est trop importante ou au contraire absente, la technique d'ensemencement a été utilisée. Celle-ci consiste à prélever des cristaux obtenus en sursaturation pour ensemencher des gouttes en sous-saturation. Les germes cristallins servant à ensemencher peuvent être soit des cristaux (on parle dans ce cas de *macroseeding*) ou alors un broyat de cristaux (*microseeding*). Dans le cas du *microseeding*, utilisé pour obtenir des cristaux de LkcE et de ces mutants, les cristaux ont été prélevés, introduit dans une solution de conservation dans laquelle la concentration en agent précipitant était légèrement

plus élevée que celle de la condition dans laquelle les cristaux ont été obtenus. Ils sont ensuite broyés à l'aide du kit Seed Bead™ (Hampton Research). Cette solution a alors été utilisée comme solution de cristallisation dans les gouttes, et une solution moins concentrée en agent précipitant a été introduite dans les puits. Le fait que la protéine n'atteigne pas la zone de nucléation, mais uniquement la zone de croissance, permet de contrôler le nombre de germes cristallins introduits par dilution du broyat.

c. Trempeage par diffusion des substrats

Dans le but d'obtenir la structure de la LkcE ou de ses mutants avec le substrat naturel de l'enzyme (LC-KA05), les cristaux ont été trempés dans une solution contenant 30 mM de substrat. En effet, la quantité de substrat produit au laboratoire et purifié par le laboratoire du Pr. Kirschning est faible, ce qui a interdit des essais de co-cristallisation. Nous avons décidé de jouer avec des durées plus ou moins longues (de 5 secondes à 20 minutes). Pour cela, le substrat a été dissous dans l'éthylène glycol, solvant qui sert également de cryoprotectant pour la congélation des cristaux. Lors du trempage, les cristaux ont été prélevés de la goutte de cristallisation puis introduits dans une solution stabilisante contenant 30 mM de LC-KA05 et 30% d'éthylène glycol. Le cristal a ensuite été congelé en flux d'azote à 100 K.

2. Congélation et enregistrement des données de diffraction

A) Congélation des cristaux

Lors de l'enregistrement des données de diffraction au synchrotron, les cristaux sont soumis à des rayons X très intenses. Ces rayonnements induisent la formation de radicaux libres (radiolyse de l'eau) ou des phénomènes d'ionisation qui entraînent la dégradation de la protéine et donc du cristal. Pour éviter les dommages dus aux radiations, et la diffusion des radicaux, il est impératif de congeler le cristal. Contrairement aux cristaux de sels, les cristaux de protéines contiennent pourcentage important de solvant (entre 30 à 70% généralement). Il est donc primordial de protéger le cristal contre la formation de glace, qui pourrait à la fois endommager le cristal, et donc affecter son pouvoir de diffraction, mais aussi conduire à la présence d'anneaux de glace sur le cliché de diffraction. Ces couronnes de glace parasites sont dues à la diffraction des rayons X par les cristaux de glace et entraîne un bruit très élevés sur les mesures des taches de diffraction à 3,5, 3,7 et 3,9 Å de résolution.

Afin de protéger les cristaux de la formation de glace, ils ont été trempés quelques secondes dans une solution stabilisante supplémentée avec 30% d'éthylène glycol. Les cristaux ont ensuite congelés sous flux d'azote à 100 K.

B) Enregistrement des données de diffraction

Les ondes diffractées par le cristal peuvent être décrites selon deux paramètres : le module $|F_{(hkl)}|$ et la phase $\Phi^{(hkl)}$. Ainsi, chaque réflexion est associée à un facteur de structure $F_{(hkl)}$ qui dépend de la nature des atomes à travers le facteur de diffusion atomique f_j et de la position (xyz) de chaque atome j au sein de la maille dans le réseau direct. Ce facteur de structure décrit ainsi la contribution de l'ensemble des atomes, à travers leurs électrons, à la diffraction selon l'Équation 1 :

$$F_{hkl} = |F_{hkl}| e^{i\varphi_{hkl}} = \sum_{j=1}^N f_j e^{[2i\pi(hx_j + ky_j + lz_j)]}$$

Équation 1 : relation de calcul du facteur de structure.

Avec $|F_{hkl}|$: amplitude ou module de l'onde diffractée, φ_{hkl} : phase de l'onde diffractée, f_j le facteur de diffusion atomique, (x_j, y_j, z_j) les coordonnées de l'atome j , et hkl les indices de Miller (ou coordonnées dans l'espace réciproque).

Il existe une relation mathématique, la transformée de Fourier, entre les amplitudes diffractées, c'est-à-dire les facteurs de structure F_{hkl} dans le réseau réciproque, et la distribution de la densité électronique du cristal $\rho(xyz)$ dans le réseau direct :

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

Équation 2 : Transformée de Fourier.

Avec (xyz) les coordonnées dans l'espace réel, hkl les indices de Miller, et V volume de la maille.

Cependant, lors de l'enregistrement d'un cliché de diffraction, l'information de phase est perdue. Seule l'amplitude de l'onde diffractée est mesurée. La phase devra donc être déterminée par une méthode indirecte (méthode développée dans le chapitre V.3.B)

Les données de diffraction ont été enregistrées au synchrotron SOLEIL (Gif-sur-Yvette) sur les lignes Proxima 1 (LkcE et mutants de LkcE) et Proxima 2 (SeLkcE).

Pour la collecte des données, le cristal est monté sur un support placé sur une tête goniométrique dans le flux d'azote et de rayons X. L'expérience consiste à irradier le cristal par les rayons X avec un temps d'exposition choisi (de 0,1 s à 0,5 s), sur un secteur angulaire dépendant du groupe d'espace, et avec un pas d'oscillations choisi par l'expérimentateur de 0,1 à 0,5° en fonction des caractéristiques propres au cristal (maille, mosaïcité, etc...).

3. Traitement des données de diffraction et phasage

A) traitement des données

Les spectres de diffraction ont été indexés et intégrés en utilisant le logiciel XDS (Kabsch, 2010) et mis à l'échelle avec XSCALE (Otwinowski, et Minor, 1997). La qualité des jeux de données enregistrées a ensuite été évaluée en fonction de plusieurs paramètres tels que la résolution maximale, la complétude, la redondance ainsi que par le facteur R_{sym} qui quantifie le désaccord entre les intensités des réflexions équivalentes par les symétries de Laue. Plus ce facteur est bas, meilleure est la cohérence interne du jeu de données. Il s'exprime par :

$$R_{\text{sym}} = \frac{\sum_{h'k'l'} \sum_{hkl} |I_{hkl} - \langle I_{h'k'l'} \rangle|}{\sum_{hkl} I_{hkl}}$$

Équation 3 : indice d'accord entre les intensités équivalentes par symétrie et leur moyenne.

$\langle I_{h'k'l'} \rangle$ est l'intensité moyenne de la réflexion unique $h'k'l'$ et I_{hkl} est l'intensité mesurée de la réflexion hkl .

B) Résolution du problème de phase

Comme nous l'avons vu précédemment, toute information de phase est perdue lors de l'acquisition des données. Pour cela, il est nécessaire d'utiliser des méthodes indirectes afin de résoudre le problème de phase.

a. Phasage par diffusion anormale

Si un cristal est centrosymétrique, la loi de Friedel indique qu'en chaque point de la maille, l'intensité d'une réflexion (h, k, l) est équivalente à celle en $(-h, -k, -l)$. Cependant, lorsqu'un atome lourd est présent au sein du cristal, et que la longueur d'onde du faisceau de rayons X est proche du seuil d'absorption de ce dernier, il se crée alors en plus un phénomène de diffusion anormale qui rompt la loi de Friedel (Blundell, et Johnson, 1976). En effet, au voisinage de ce seuil, il y a un phénomène de résonance entre les photons X et le nuage électronique de l'atome anormal. Il en découle un déphasage entre l'onde incidente et l'onde réémise par l'atome. Le facteur de diffusion atomique s'écrira alors $f(\lambda) = f^{\circ} + f'(\lambda) + if''(\lambda)$ avec f° le facteur de diffusion en absence de diffusion anormale, $f'(\lambda)$ la partie anormale réelle (dispersive) et $f''(\lambda)$ la partie anormale imaginaire due à l'absorption des rayons X par les électrons du diffuseur anormal.

Il existe deux méthodes de phasage utilisant la diffusion anormale : le SAD (*single wavelength anomalous dispersion*) et le MAD (*multiple wavelength anomalous diffraction*). Le principe des deux méthodes consiste à déterminer la position de l'atome lourd dans la maille cristalline.

i. La diffusion anormale simple (SAD)

En SAD, une seule longueur d'onde est utilisée au cours de l'enregistrement des données. Expérimentalement, nous effectuons un scan de longueur d'ondes pour déterminer le pic d'absorbance du diffuseur anormal (Figure 62). Les données sont alors enregistrées à la longueur d'onde du maximum du pic d'absorbance du diffuseur anormal. Afin de pouvoir obtenir les phases expérimentales, il est nécessaire de localiser la position des atomes lourds dans la maille du cristal. Celle-ci peut être déterminée par la fonction de Patterson (Patterson, 1935). Étant donné que les amplitudes des deux contributions réelle (f') et imaginaire (f'') de l'atome lourd sont

connues, il est maintenant possible de déterminer le facteur de structure grâce à la méthode de Harker (Figure 63). Cependant, celle-ci donne une ambiguïté sur la phase.

Deux solutions sont possibles car la méthode de Patterson ne permet pas de distinguer les positions énantiomériques des atomes lourds. Il y a donc deux cartes de densité différentes équiprobables. La moyenne des deux solutions donne naissance à une carte de Fourier qui est la superposition de deux cartes : l'une représentant la densité électronique de la protéine et l'autre représentant du bruit attribuable à aucun atome particulier. Afin d'améliorer cette carte, et donc de lever l'ambiguïté, une procédure itérative d'amélioration de densité est effectuée en utilisant des algorithmes comme l'aplatissement de solvant (Wang, 1985) et en effectuant la moyenne des symétries non-cristallographique (NCS) à l'aide du programme PARROT (Zhang, *et al.*, 1997).

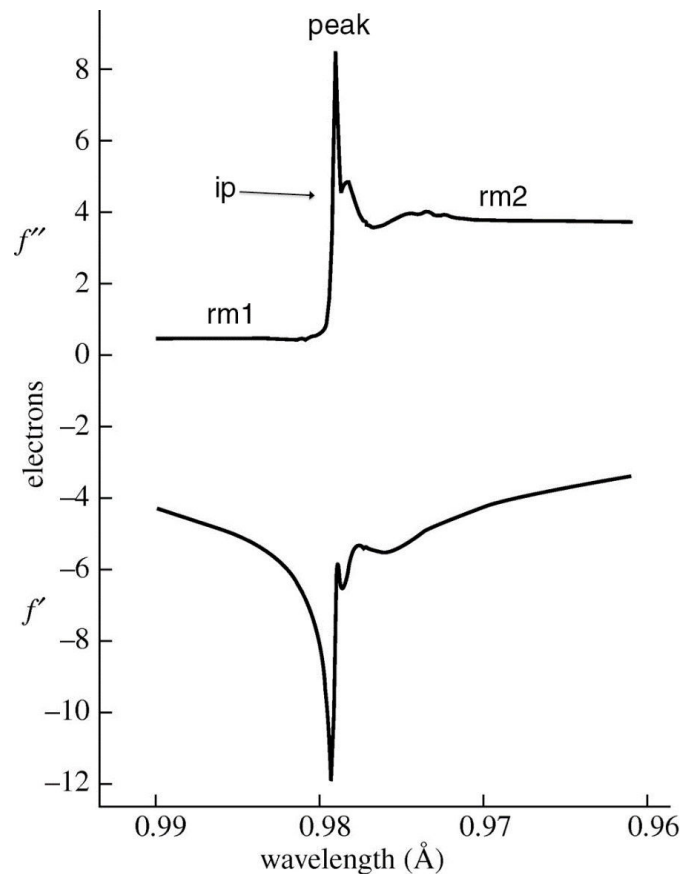


Figure 62 : Scan de fluorescence du sélénium sur une protéine sélénée (adapté de Duke, et Johnson, 2010).

A proximité du seuil d'absorption les parties réelle et imaginaire f' et f'' varient rapidement. Pour un jeu SAD, seule une acquisition à une longueur d'onde du pic d'absorbance du diffuseur anomal (peak ou pk) est enregistrée. Pour l'acquisition d'un jeu MAD, les longueurs d'onde du pic, du point d'inflexion (ip), avant le pic d'absorbance (rm1) et après celui-ci (rm2) sont utilisés.

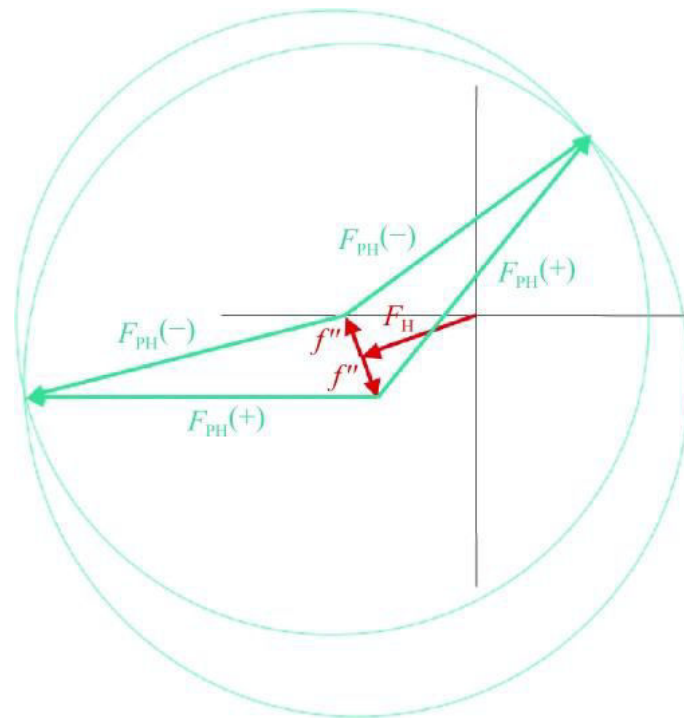


Figure 63 : Méthode de Harker permettant de déterminer la phase par SAD (tiré de Taylor, 2010).

Cette représentation montre que la méthode SAD induit une ambiguïté de phase dans lesquels deux solutions (appelées « mains ») sont possibles, solutions qui correspondent aux intersections entre les deux cercles.

ii. Le MAD

Pour un phasage par MAD, plusieurs jeux de données sont nécessaires. Bien qu'en théorie deux longueurs d'ondes soient nécessaires pour le phasage en MAD, trois, voire quatre jeux de données à des longueurs d'onde différentes sont enregistrés. Un jeu de données est enregistré sur le pic d'absorbance du diffuseur anormal, un sur le point d'inflexion de la courbe, et un ou deux éloignés du pic d'absorbance (*remote 1* et *remote 2*) (Figure 62). Ces jeux de données sont enregistrés de préférence sur un même cristal pour éviter des problèmes de non-isomorphisme. Contrairement au SAD, le fait d'avoir un jeu de données à plusieurs longueurs d'onde permet de lever l'ambiguïté de phase.

iii. Phasage expérimental de la LkcE

Bien qu'un jeu MAD ait été enregistré sur la ligne de lumière Proxima 2, le phasage pour la SeLkcE a été effectué par SAD sur le jeu de données du *peak*, enregistrés à une longueur d'onde de 0,9793 Å. Pour cela, nous avons utilisé le logiciel HKL2MAP (Pape, et Schneider, 2004) qui est une interface graphique employant la suite SHELX (Sheldrick, 2010). Les données de diffusion anormale ont tout d'abord été traitées avec XPREP. Ce logiciel permet de définir la résolution maximale du jeu de données anormales à 3,3 Å, et d'estimer les facteurs de structure anormale. La suite SHELX a ensuite permis d'ajouter et d'affiner la position de 23 atomes de sélénium dans la maille. Parmi ces 23 atomes, 16 ont un taux d'occupation supérieur à 0,7. Ainsi les sites de 16 des 18 atomes de sélénium incorporés dans la SeLkcE ont pu être déterminés sans ambiguïté. La figure de mérite (FOM) sur le positionnement des atomes lourds était de 0,376.

Afin de discerner entre les deux mains possibles nous avons utilisé le logiciel PARROT (Zhang, *et al.*, 1997) pour effectuer un aplatissement de solvant. Le programme utilise également les symétries non cristallographiques (NCS) déterminées au préalable sur la position des atomes lourds par SHELX. PARROT, en plus de sélectionner la bonne solution de phasage, a amélioré considérablement les phases, ce qui se traduit par une figure de mérite sur les phases qui augmente de 0,374 à 0,867 et une corrélation de 0,690 qui augmente jusqu'à 0,939 au terme de 21 cycles. Ceci a permis d'obtenir une carte exploitable dans laquelle les structures secondaires étaient clairement identifiables. Cette carte de densité a permis la construction de la première structure pour la LkcE.

b. Phasage par remplacement moléculaire

La résolution du problème de phase en utilisant le remplacement moléculaire est la technique la plus utilisée à l'heure actuelle. En effet, cette technique nécessite un modèle tridimensionnelle suffisamment proche de la structure de la protéine étudiée pour pouvoir résoudre le problème de phase (Rossmann, 1990). Etant donné le nombre croissant de structures au sein de la Protein Data Bank (PDB) (>111 000), nous pouvons aisément comprendre l'intérêt croissant de cette méthode de phasage simple et rapide à mettre en œuvre. Pour phaser par remplacement moléculaire, il est conseillé de disposer d'un modèle présentant au minimum 30% d'identité de

séquence. Le but du remplacement moléculaire est de positionner une ou plusieurs molécules modèles dans la maille cristalline afin de pouvoir calculer une première estimation des phases pour les facteurs de structure. Il faut pour cela déterminer les fonctions de rotation et de translation permettant de positionner au mieux le modèle pour simuler le contenu de la maille (Rossmann, et Blow, 1962).

Nous avons utilisé Phaser MR (McCoy, *et al.*, 2007) de la suite CCP4 (Winn, *et al.*, 2011) pour résoudre le problème de phase des jeux de données enregistrés sur les protéines natives en utilisant le modèle construit à partir du modèle de la protéine sélénée. Ce programme utilise le maximum de vraisemblance afin de trouver la meilleure solution pour les fonctions de rotation et de translation. Tout d'abord, le modèle est orienté par la recherche d'une solution de rotation, puis sa position dans l'unité asymétrique est retrouvée par une étape de translation. La LkcE étant un homodimère, le remplacement moléculaire a été effectué par une recherche à corps rigide de deux copies du monomère reliés par une NCS.

4. Construction et affinement des structures

Après avoir résolu la phase par SAD, un modèle pour la SeLkcE a été construit à l'aide du programme Buccaneer (Cowtan, et Zhang, 1999). Ce programme est particulièrement adapté pour construire un modèle protéique à partir de la séquence de la protéine au sein d'une densité électronique à une faible résolution (>2,5 Å). Buccaneer construit un modèle partiel de la protéine qui est ensuite complété et modifié manuellement à l'aide du logiciel Coot (Emsley, *et al.*, 2010). Deux tours d'affinement ont été réalisés à l'aide du programme REFMAC (Murshudov, *et al.*, 1997).

Après phasage par remplacement moléculaire à partir du modèle construit pour la SeLkcE, la structure de la LkcE native a été construite à l'aide du logiciel Coot (Emsley, *et al.*, 2010). Les phases ont ensuite été affinées en utilisant le programme BUSTER (Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2016). BUSTER version X.Y.Z. Cambridge, United Kingdom: Global Phasing Ltd). Nous avons utilisé ce programme d'affinement car il présente l'avantage d'être particulièrement efficace sur les basses résolutions et d'améliorer la densité électronique des boucles flexibles. De la même façon, BUSTER améliore également la densité correspondant aux analogues de substrat ou au substrat naturel, dans les cas de co-cristallisation ou de trempages. Le

but de l'étape d'affinement est de réduire l'écart entre les facteurs de structures observés (F_{obs}) et ceux calculés à partir du modèle (F_{calc}). Ce processus est réalisé par une alternance de construction manuelle du modèle dans Coot et des cycles d'affinements numériques des paramètres des atomes. Ainsi des contraintes d'ordre stéréochimique, telles que la longueur et les angles de liaisons, la gestion des clashes stériques, et d'ordre cristallographique tels que les NCS vont être apportés au modèle.

Le modèle construit a été évalué à chaque cycle d'affinement. Pour cela, le facteur R_{factor} permet de juger de l'accord entre le modèle et les facteurs de structure observés. Un facteur R_{free} , calculé à partir d'un échantillon de réflexions (5%) non utilisées au cours de l'affinement, permet également d'éviter une surinterprétation de la carte de densité. Ces deux facteurs doivent être le plus faible possible et l'écart entre les deux facteurs doit rester de l'ordre de 5%.

La qualité de la structure est enfin vérifiée grâce au serveur MolProbity (Davis, *et al.*, 2007 ; Chen, *et al.*, 2010). Ce serveur permet la validation de la structure en vérifiant la géométrie de la protéine, le respect de la carte de Ramachandran et les clashes stériques, et propose un score pour juger de la qualité du modèle.

VI. Test d'activité et détermination des paramètres cinétiques de la LkcE

1. Test d'activité de la LkcE avec analogues de substrat

Afin de tester l'activité de LkcE, ainsi que la co-cristallisation, nous disposons de deux analogues de substrat mimant chacun une partie du substrat naturel de l'enzyme (Figure 64).

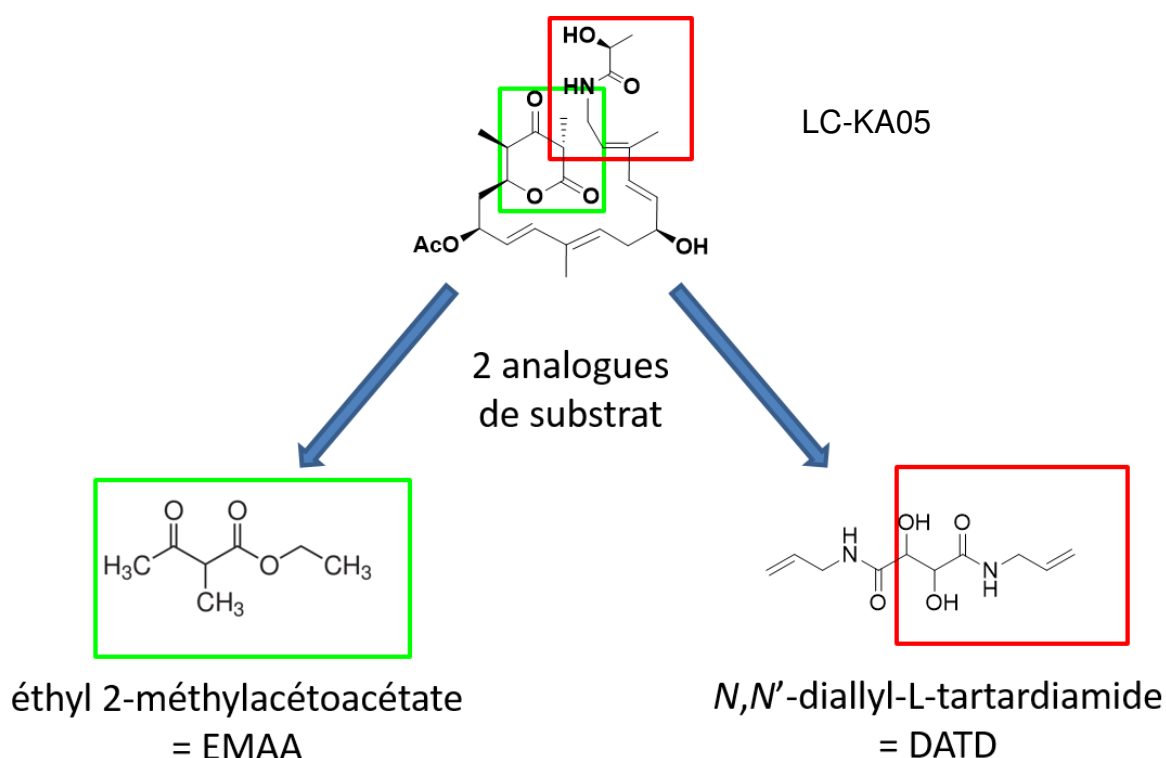


Figure 64 : Analogues du substrat naturel de la LkcE.

L'éthyl 2-méthylacétoacétate (EMAA) mime la partie lactone du LC-KA05 (en vert), tandis que le *N,N'*-diallyl-L-tartardiamide (DATD) mime la partie amide (en rouge).

La LkcE (50 μ M) a été incubée dans différentes concentrations des deux analogues de substrats. Le mélange réactionnel a été placé à 25 °C durant 1 h. 0,1% d'acide trichloroacétique (TCA) a été ajouté pour stopper la réaction. Après centrifugation 5 min à 15 000 *g*, 100 μ L ont alors été injectés sur colonne C8 Vydac 208TP52, 2.1 x 250 mm, 5 μ m (PerkinElmer Life Sciences). Les molécules ont alors été éluées par un gradient croissant d'acétonitrile (de 0% jusqu'à 80%).

La réaction a également été reproduite de la même façon, mais a cette fois-ci, été arrêtée avec ajout d'acétate d'éthyle (deux fois un équivalent volume). La phase organique a alors été récupérée analysés en LC-MS (chromatographie phase inverse couplée à la spectrométrie de masse) par C. Paris (LiBio, Université de Lorraine).

2. Test d'activité de la LkcE avec son substrat naturel (LC-KA05)

La LkcE (50 μM ou 100 μM) a été mise en présence de concentrations différentes en LC-KA05 (10 à 250 μM) et placée à 25 °C. La réaction a été arrêtée à différents temps en prélevant 100 μL du mélange réactionnel et en le mélangeant à 100 μL d'acétate d'éthyle. La phase organique a été conservée. La phase organique a alors été extraite une seconde fois avec 100 μL d'acétate d'éthyle. Les phases organiques des deux réactions ont été mises en commun. Le solvant a alors été évaporé sous pression réduite. 100 μL d'un mélange eau : acétonitrile (20:80) a été utilisé pour reprendre les molécules avant injection en LC-MS, analyses réalisées en collaboration avec C. Paris (LiBio, Université de Lorraine).

Au préalable, une gamme étalon a été établie de la même façon, en utilisant des concentrations allant de 5 à 250 μM , mais en absence d'enzyme. Cette gamme étalon a ensuite permis de quantifier la quantité de substrat transformée par l'enzyme en reportant les valeurs d'intensité obtenue pour chacun des tests. .

VII. La diffusion des rayons X aux petits angles (SAXS)

1. Principe et acquisition des données

Le SAXS permet de déterminer la masse moléculaire, les dimensions et la forme globale d'une protéine en solution. Cette technique d'étude structurale est parfaitement adaptée à l'étude d'objets qui présentent une certaine flexibilité en solution, ce qui rend leur cristallisation très difficile. Le SAXS constitue de cette façon une approche complémentaire à la cristallographie, mais aussi à la RMN puisqu'il n'y a pas de limitation de taille pour l'objet étudié. Les données enregistrées sont en revanche à faible résolution. Il est difficile d'estimer une résolution en SAXS, puisque cette méthode permet de déterminer l'enveloppe globale de la protéine, mais aussi la

construction de modèles tout atome à partir des structures à haute résolution, mais il est généralement admis que les données enregistrées présentent une résolution limite supérieure à la dizaine d'Angström (Rambo, & Tainer, 2013).

Le SAXS, tout comme la cristallographie, est basé sur l'interaction des rayons X avec les électrons de la protéine. Le faisceau de rayons X, de longueur d'onde λ , va alors être diffusé au contact de la protéine, et les rayons diffusés aux petits angles ($0,1 - 3^\circ$) vont être analysés sur un détecteur. Les données diffusées à angle nul sont en revanche non mesurables dû à la présence d'un *beam stop* pour protéger le détecteur du faisceau direct. Nous avons effectué nos mesures sur la ligne de lumière SWING du synchrotron SOLEIL, qui permet de réaliser des mesures de SEC-SAXS (SAXS couplé à une chromatographie d'exclusion stérique en HPLC) (Putnam, *et al.*, 2007 ; Pérez, & Koutsioubas, 2015) (Figure 65). Ceci permet une dernière purification de l'échantillon avant la mesure, et d'éviter une contamination par des agrégats éventuels, ou de faire des mesures sur des complexes protéiques en séparant le complexe des protéines dissociées. La présence d'agrégats durant la mesure serait en effet délétère étant donné que la diffusion des particules en solution est proportionnelle à la masse des solutés au carré.

Les données expérimentales consistent en une courbe d'intensité (I) en fonction du vecteur de diffusion q ($q = \frac{4\pi \sin\theta}{\lambda}$), avec 2θ l'angle entre le faisceau direct et le rayon diffusé ; q a les dimensions de l'inverse d'une distance, et est exprimé généralement en \AA^{-1} ou nm^{-1} (Figure 66)

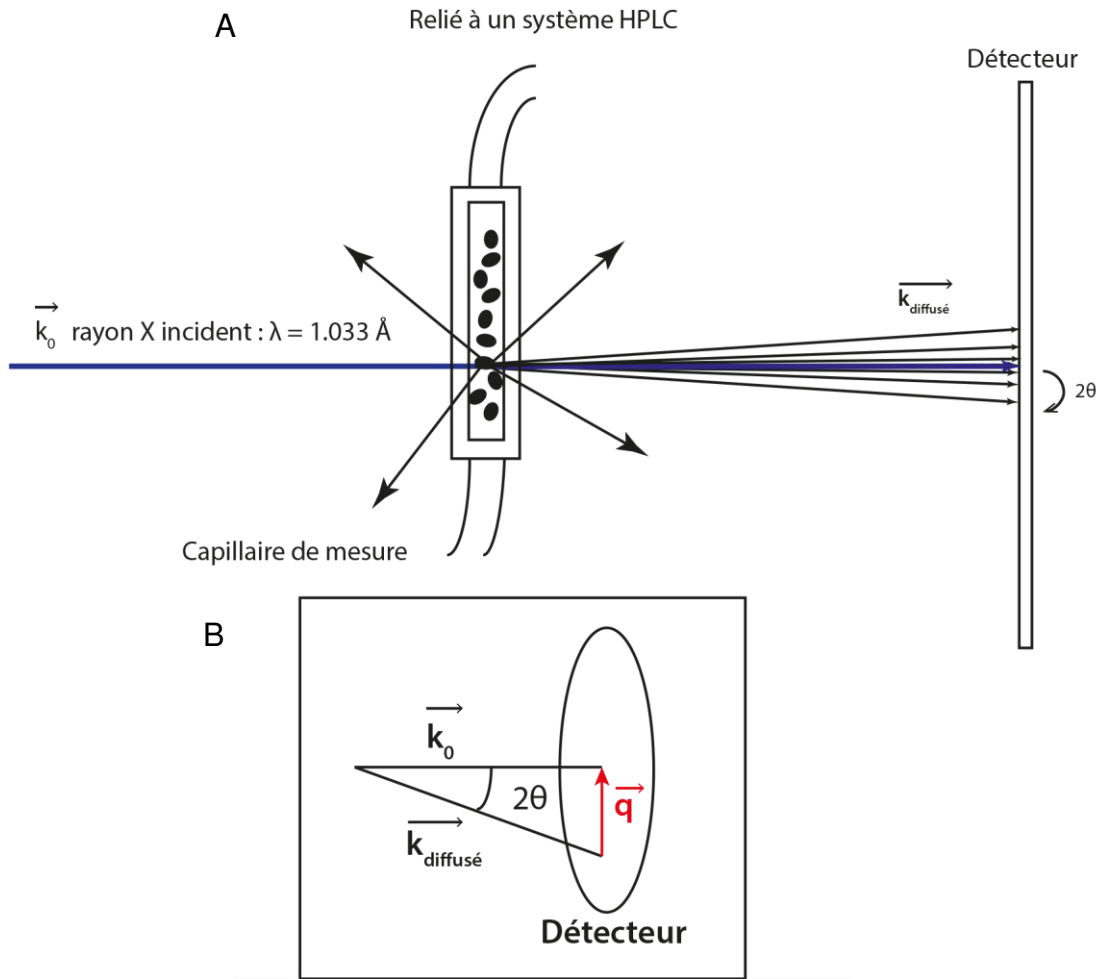


Figure 65 : Description du principe du SAXS.

A : Sur la ligne SWING de SOLEIL, le capillaire de mesure est couplé à une colonne de type SEC 3 ou 5 (Agilent) couplée à un système HPLC, ce qui permet de séparer les agrégats de la protéine d'intérêt. Le faisceau X incident k_0 (en violet) interagit avec l'échantillon protéique. L'échantillon va alors diffuser les rayons X dans toutes les directions. Le détecteur est placé à une distance variant de 1 à 4 m de l'échantillon pour enregistrer le rayonnement diffusé aux petits angles ($0,1^\circ$ à 3°). Un "beam stop" en amont du détecteur permet d'arrêter le faisceau incident direct de trop forte intensité.

B : Le vecteur q de diffusion est tel que : $\vec{q} = \vec{k}_{\text{diffusé}} - \vec{k}_0$, et dont la longueur vaut $q = 4\pi\sin\theta/\lambda$.

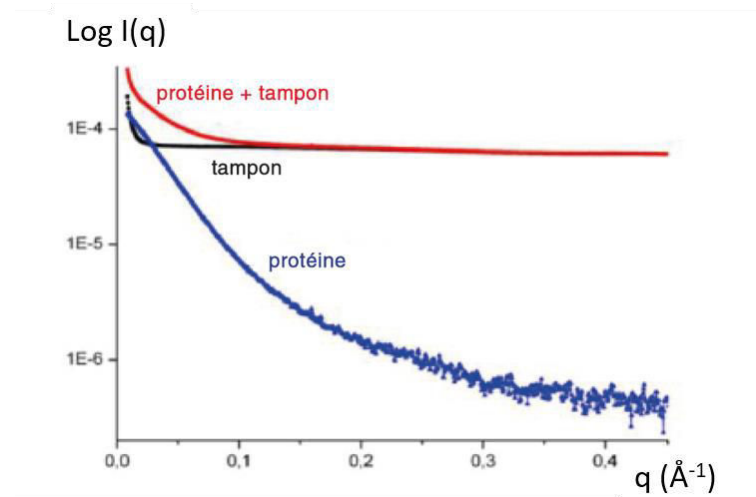


Figure 66 : Description de l'enregistrement des données SAXS.

Dans un premier temps, une série de 120 images de diffusion du tampon est enregistrée, puis dans un second temps, une série de 240 images de diffusion entourant le pic d'éluion de la protéine dans le tampon est enregistrée. Le temps d'acquisition de chaque image est de 750 ms avec un temps de latence entre chaque image de 250 ms. La moyenne radiale des intensités diffusées sur le détecteur est calculée pour chaque image, avec le logiciel Foxtrot (Girardot, *et al.*, 2011). Les courbes de diffusion du tampon sont ensuite moyennées, puis soustraites aux courbes de diffusion de la protéine dans le tampon pour obtenir uniquement le signal de diffusion de la protéine seule.

1. Traitement des données

A) Loi et Représentation de Guinier

La loi de Guinier permet d'approximer l'intensité $I(q)$ par la relation : $\ln I(q) \approx \ln I(0) - q^2 \frac{R_g^2}{3}$. Elle n'est valable qu'aux très petits angles (pour des valeurs de q faibles), définis pour des valeurs telles que $qR_g < 1,3$. L'application de cette loi, représentant $\ln I(q)$ en fonction de q^2 , nous permet ainsi de déterminer le rayon de giration R_g de la protéine à partir de la pente de la droite ainsi que d'extrapoler l'intensité diffusée à angle nul $I(0)$. Cette représentation, appelée représentation de Guinier, nous permet d'autre part de déterminer la qualité des mesures et de l'échantillon, la linéarité étant un gage de la qualité des données et la preuve de l'absence d'interactions intermoléculaires (Figure 67).

Sur la ligne de lumière, cette représentation de Guinier nous permet de calculer le R_g et le $I(0)$ pour chaque image enregistrée. Ainsi seules les images consécutives ayant un R_g constant et une valeur de $I(0)$ assez élevée seront moyennées. En effet, la valeur de $I(0)$ dépend de la concentration protéique et des valeurs élevées

permettent ainsi d'augmenter le ratio signal sur bruit. Cependant, cette valeur est également corrélée au carré de la masse moléculaire de l'espèce présente en solution et des agrégats vont donc présenter une valeur de $I(0)$ importante. Ceci nous permet de sélectionner les images enregistrées sur une seule et même espèce protéique, et de moyennner celles-ci afin d'augmenter le ratio signal/bruit (Figure 68).

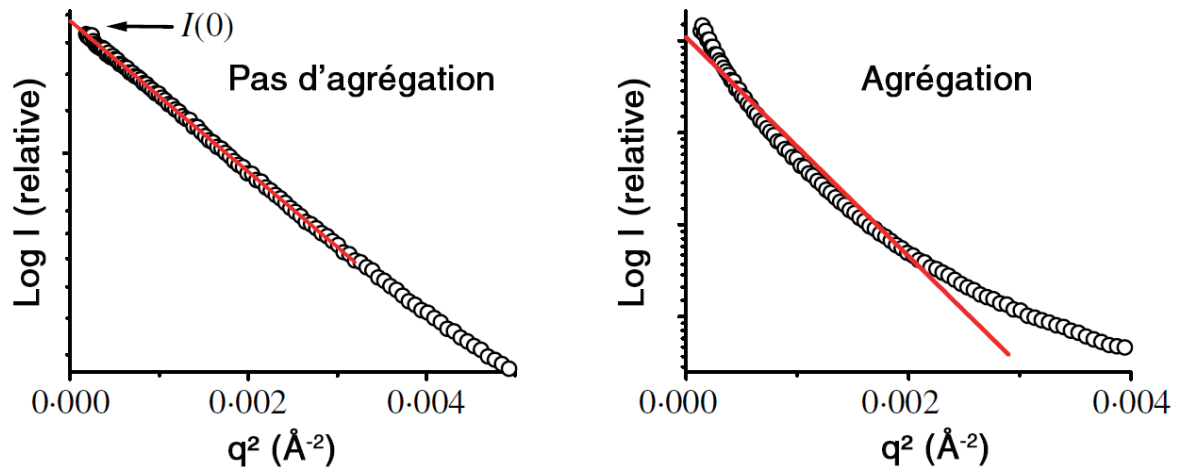


Figure 67 : Représentation de Guinier des données SAXS (adapté de Putnam, et al., 2007).

La loi de Guinier est valable pour des valeurs de $qR_g < 1,3$. Elle se calcule donc aux très petits angles, ce qui permet d'étudier les dimensions de la particule en solution mais également les interactions intermoléculaires. Ainsi, lorsque dans cette représentation, l'intensité diffusée forme une droite, la solution étudiée se rapproche d'une solution idéale et les données sont alors exploitables. Cette représentation peut traduire des interactions non souhaitables au sein de la solution étudiée, puisque la courbe peut alors présenter une intensité croissante aux faibles valeurs de q^2 (agrégats ou interactions attractives) ou au contraire, des valeurs d'intensités décroissantes (interactions répulsives). La représentation de Guinier permet aussi d'extrapoler la valeur de $I(0)$ nécessaire pour calculer la masse moléculaire de la protéine en solution.

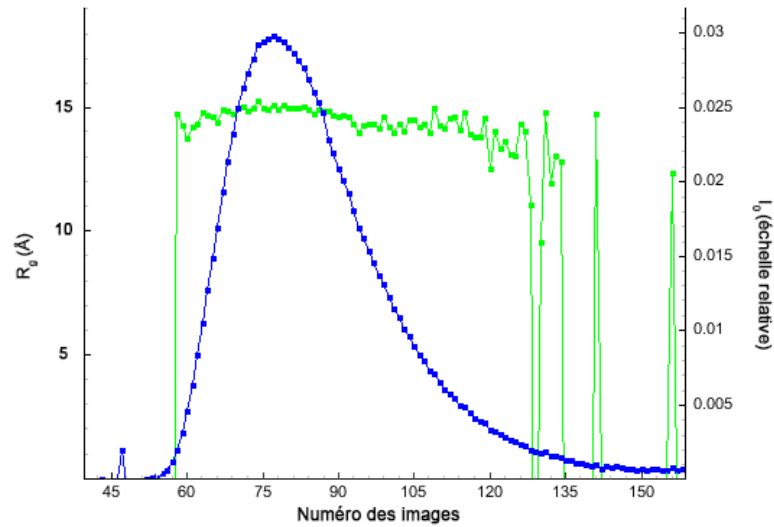


Figure 68 Représentation du rayon de gyration et de l'intensité de diffusion à angle nul en fonction des images enregistrées.

Une représentation de Guinier est effectuée grâce au programme Foxtrot sur chaque image enregistrée sur la protéine. Seules les images pour lesquelles le R_g (en vert) est constant sont moyennées. La valeur de $I(0)$ (en bleu) reflète, quant à elle la concentration protéique et donc, le pic d'élution en HPLC.

B) Fonction de distribution de distance

Cette fonction est calculée grâce à la transformée de Fourier de la courbe de diffusion : $P(r) = F(I(q)) = \int I(q) e^{-i\vec{q}\vec{r}} dq$. Elle constitue un histogramme des distances inter-électroniques au sein de la protéine dans l'espace réel. Elle représente ainsi la probabilité de trouver une distance entre deux électrons au sein de l'objet étudié en solution, en fonction de la distance elle-même (Figure 69). Le calcul de la fonction de distribution de distance se fait *via* le programme GNOM de la suite ATSAS (Svergun, 1992). Cette fonction permet de définir le D_{\max} , c'est-à-dire la distance maximale entre deux électrons au sein de la protéine étudiée. En effet, au-delà de cette valeur, la probabilité de trouver une distance entre deux électrons devient nulle. L'allure de la fonction de distribution de distance renseigne également sur la forme de l'objet en solution. La fonction de distribution des distances constitue une deuxième méthode de détermination du R_g . En effet, le R_g peut être déterminé à partir de l'équation suivante : $R_g^2 = \int_0^{D_{\max}} r^2 P(r) dr / \int_0^{D_{\max}} P(r) dr$. La valeur de $I(0)$ peut également être déterminée à partir de la fonction de distribution de distances : $I(0) = 4\pi \int_0^{D_{\max}} P(r) dr$. Ces valeurs de R_g et $I(0)$ sont ainsi déterminées sur toute la gamme

de q (contrairement à la loi de Guinier). Ceci permet d'atténuer la contribution des interactions attractives ou répulsives dans la détermination de ces deux valeurs. Il s'agit alors d'une méthode efficace pour confirmer les valeurs de R_g et de $I(0)$ déterminées par la loi de Guinier.

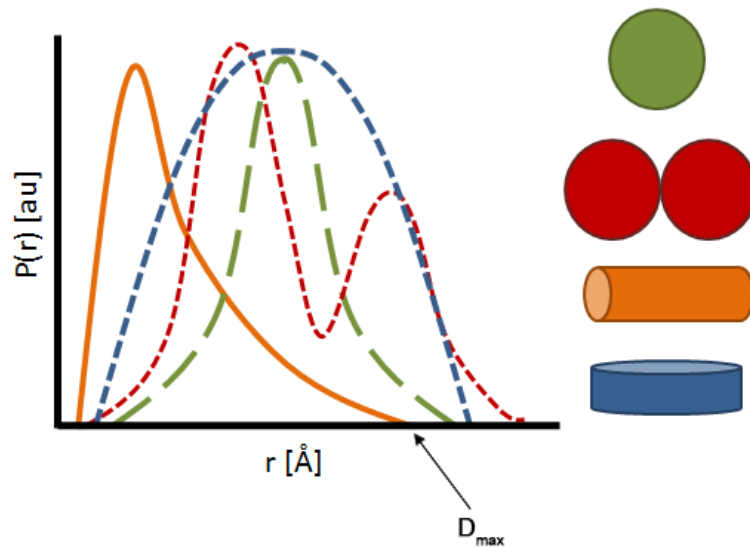


Figure 69 : Fonction de distribution de distance (tiré de Nyman, & Fullmer, 2015)

Plusieurs fonctions de distribution de distance sont représentées ici. Chaque courbe correspond à une des formes représentées sur la droite. Le D_{max} pour la courbe en orange représentant un cylindre est indiqué à titre d'exemple.

C) Représentation de Kratky

Le graphique de Kratky est une façon particulière de représenter les données SAXS. Elle représente $q^2 I(q)$ en fonction de q (Kratky, & Porod, 1949). Cette fonction peut ensuite être normalisée en multipliant le vecteur d'onde q par le R_g , ce qui permet de rendre l'angle de diffusion indépendant des dimensions de la protéine étudiée, et $I(q)$ peut être divisée par le $I(0)$ afin d'affranchir les valeurs d'intensité de la masse moléculaire de la protéine (Durand, *et al.*, 2010). Ceci permet de comparer directement les représentations de Kratky pour des protéines différentes.

Cette représentation fournit des informations sur le niveau de structuration de la protéine. En effet, si cette représentation forme une courbe en cloche, cela signifie que la protéine est globulaire en solution. En revanche, si cette fonction montre un plateau,

ceci est typique d'une protéine non globulaire, étendue et probablement flexible (Figure 70).

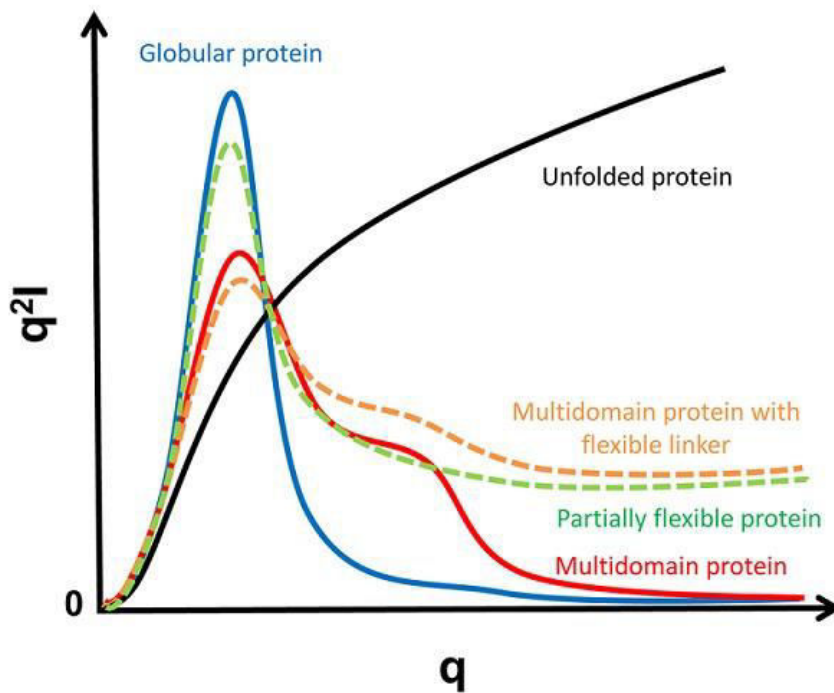


Figure 70 : Représentation schématique de Kratky obtenue sur divers types de protéine (figure réalisée par SSRL, Stanford).

L'allure de cette représentation dépend du caractère ordonnée ou non de la protéine, de sa composition en domaine et de la flexibilité de la protéine ou des linkers la composant.

2. Interprétation des données SAXS et modélisation structurale

Plusieurs programmes, majoritairement présents dans la suite ATSAS, sont ensuite utilisés pour interpréter les données SAXS. Ces programmes cherchent généralement à corrélérer les courbes de diffusion calculées pour les modèles qu'ils proposent, à la courbe de diffusion expérimentale. Pour cela, un facteur d'accord est alors calculé : le chi2 (χ^2). Ce test du chi2, bien qu'étant représentatif de l'accord entre les données calculées et les données expérimentales (comme peuvent l'être les facteurs R pour la cristallographie) présente certaines limites. En effet la valeur de χ^2 est forcément sensible à l'erreur sur la mesure de la courbe SAXS. Or il se trouve que sur certaines lignes de lumière, utilisant en particulier la technique du SEC-SAXS, ces erreurs sur la mesure sont sous-déterminées. D'autres méthodes pour juger de la qualité des modèles proposés sont actuellement en développement, telle que

l'utilisation de cartes de corrélation (Franke, *et al.*, 2015). Je me limiterai ici à une brève description des différents programmes utilisés dans les publications jointes à cette thèse.

A) Calcul d'enveloppe *ab-initio*

Le SAXS permet de déterminer l'enveloppe de la protéine à basse résolution, et ce, grâce notamment à deux programmes majoritairement utilisés : DAMMIN (Svergun, 1999) et GASBOR (Svergun, *et al.*, 2001).

DAMMIN crée un modèle à partir de pseudo-atomes de rayon variable (typiquement de 3 Å à 10 Å). Cette valeur dépend de la taille de la protéine étudiée, le programme utilisant environ 1500 pseudo-atomes par unité asymétrique pour le volume étudié. A partir d'une forme compacte de pseudo-atomes définie par l'utilisateur (ellipse, sphère, etc...) d'un diamètre correspondant au D_{max} de la protéine, le programme va procéder à des cycles itératifs de recuit simulé. A chaque cycle, DAMMIN va supprimer des pseudo-atomes et/ou les réarranger dans l'espace, en gardant comme contrainte la valeur de D_{max} . De la même façon, afin de garantir la cohérence du calcul d'enveloppe, il applique une pénalité de score lorsque les pseudo-atomes sont discontinus. Il va alors calculer une courbe de diffusion théorique à partir du modèle à chaque cycle et va tenter de la faire correspondre au mieux à la courbe de diffusion expérimentale. DAMMIN utilise seulement une partie de la courbe de diffusion allant jusque des valeurs de q de $0,25 \text{ \AA}^{-1}$. Le programme MONSA permet quant à lui de réaliser des calculs d'enveloppe en utilisant le même algorithme que DAMMIN mais il peut prendre en compte plusieurs courbes de diffusion différentes de façon simultanée. Son utilisation permet ainsi de discriminer la position d'un sous-ensemble (domaine, etc...) au sein de la protéine entière ou d'un complexe.

GASBOR se base sur les mêmes algorithmes que DAMMIN, mais utilise quant à lui des pseudo-résidus au lieu de pseudo-atomes, dont le diamètre correspond à la distance entre deux acides aminés consécutifs, soit 3,8 Å. Ainsi l'utilisateur détermine le nombre de pseudo-résidus utilisés, qui doit être équivalent au nombre de résidus présents dans la protéine étudiée. GASBOR présente aussi l'avantage de prendre en compte la totalité de la courbe de diffusion qui lui est fournie. Il permet ainsi de réaliser des calculs d'enveloppe à plus haute résolution que DAMMIN, allant ainsi jusque 10 Å de résolution dans la pratique et jusque $1,2 \text{ \AA}^{-1}$, soit 5 Å de résolution dans la théorie

(les données SAXS sur des échantillons biologiques sont rarement enregistrées à plus de $0,6 \text{ \AA}^{-1}$).

Pour chaque construction, un minimum de dix calculs de formes sont effectués. Les enveloppes similaires sont ensuite superposées par DAMSUP/SUPCOMB, moyennées par DAMAVER puis filtrées par DAMFILT afin de ne conserver que des portions qui sont redondantes entre plusieurs enveloppes mais aussi d'éliminer les pseudo-atomes faiblement connectés au reste de l'enveloppe afin de rendre celle-ci plus compacte (Volkov, & Svergun, 2003) (Figure 71). DAMSTART peut également être utilisé suite à DAMAVER pour générer un ensemble servant de point de départ à DAMMIN, ce qui permet d'affiner le calcul d'enveloppe.

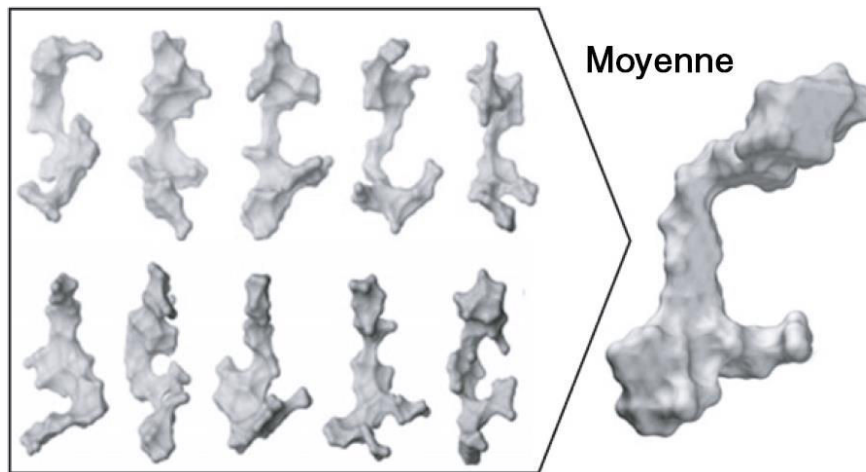


Figure 71 : Illustration du processus de calcul d'enveloppe ab-initio (tiré de Putnam, et al., 2007).

Dix calculs d'enveloppe sont réalisés avec DAMMIN ou GASBOR puis moyennés en utilisant la suite DAMFILT. L'enveloppe obtenue traduit une moyenne des conformations correspondant aux données SAXS.

B) Comparaison de structures tout atome aux données SAXS

Un des points très intéressants du SAXS est de pouvoir comparer les structures à haute résolution résolues par cristallographie aux données en solution. En effet, il arrive parfois que les structures cristallographiques de certaines protéines ne reflètent pas leur état physiologique en solution. Les contacts avec les symétriques induits par l'empilement cristallin peuvent en effet favoriser certaines conformations alors qu'elles ne sont pas majoritaires lorsque la protéine est en solution.

Pour cela le programme CRY SOL (Svergun, *et al.*, 1995) calcule une courbe de diffusion théorique pour la structure à haute résolution et la superpose ensuite à la courbe de diffusion expérimentale. Si un désaccord apparaît, des programmes de dynamique moléculaire ou d'utilisation de mode normaux peuvent être utilisés pour générer de la diversité dans les modèles, permettant ainsi d'améliorer l'accord entre les données calculées et expérimentales (Suhre, & Sanejouand, 2004 ; Panjkovich, & Svergun, 2016).

Dans le cas d'une hétérogénéité en solution, que ce soit au niveau oligomérique ou due à des changements conformationnels important, le programme OLIGOMER peut calculer une courbe théorique d'un ensemble de conformères et l'ajuster à la courbe expérimentale. Il permet ainsi de déterminer la proportion des différentes espèces présentes en solution.

C) Modélisation en corps rigide

SASREF (Petoukhov, & Svergun, 2005) est un programme de modélisation de structures quaternaires. En effet, il s'agit d'un logiciel de recuit simulé en corps rigide. A partir de plusieurs fichiers PDB source, il va les réorganiser dans l'espace et calculer à chaque cycle une courbe de diffusion de l'ensemble afin de la faire correspondre aux données SAXS expérimentales. Il va également éviter les clashes stériques entre les différents corps rigides. Il est possible de lui imposer des distances maximales entre résidus de fichiers PDB différents, ce qui permet à la fois de positionner et d'orienter initialement les différentes structures tout atome afin d'aider le programme à converger. Il est ainsi possible de positionner les structures à haute résolution de partenaires lors de l'étude d'un complexe ou les différents domaines entre eux lors de l'étude de systèmes multi-domaines.

D) Détermination de la masse moléculaire sans utilisation du $I(0)$

Déterminer la masse moléculaire de la protéine en solution peut être nécessaire pour connaître l'état oligomérique de la protéine lors de l'enregistrement des données ou vérifier si un complexe est bien formé. Une des méthodes consiste à enregistrer au moins deux jeux de données à des concentrations différentes (Plestil, *et al.*, 1991). Une autre technique consiste à utiliser une protéine standard de concentration connue (Mylonas, & Svergun, 2007) comme référence, ou à comparer les intensités absolues mesurées à celle de l'eau pure. Cependant ces méthodes de détermination de la masse moléculaire nécessitent de connaître la concentration de l'échantillon protéique étudiée. Or, lors de l'utilisation de la technique de SEC-SAXS, le passage préalable de la protéine sur colonne permet difficilement de déterminer la concentration exacte de la protéine.

Le programme SAXS-MoW (Fischer, *et al.*, 2010) utilise, quant à lui, une autre méthode, indépendante de la concentration de la solution en protéine. En effet, ce programme essaie tout d'abord de trouver le meilleur Guinier afin de déterminer le plus précisément possible la valeur de $I(0)$. Dans la première version du logiciel, cette valeur de $I(0)$ était déterminée grâce à GNOM, à partir du tracé de la fonction de distribution de distance. Ensuite, le programme calcule l'invariant de Porod Q :

$$Q = \int_0^{\infty} I(q)q^2 dq = 2\pi^2(\Delta\rho)^2NV$$

où N est le nombre de molécules de protéine par unité de volume, V est le volume de la protéine, et $\Delta\rho$ la différence entre la densité électronique moyenne de la protéine et celle du tampon. Ceci va ainsi permettre de déterminer le volume de la protéine $V = 2\pi^2 \frac{I(0)}{Q}$ et d'en déduire sa masse moléculaire via sa densité.

RESULTATS

I. Structure en solution d'un module de PKS *trans*-AT : un premier pas dans la compréhension de la communication intra-modulaire

Ce travail a fait l'objet d'une publication dans laquelle je suis second auteur :

Davison, J., **Dorival, J.**, Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A. et 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).

Cette étude, menée sur le module 5 de la PKS synthétisant la virginiamycine M, visait à la fois à tenter d'élucider le fonctionnement des tandems d'ACP, mais surtout de comprendre les mécanismes d'interaction entre domaines au sein d'un module PKS puisque aucune donnée structurale sur un module entier n'était alors disponible. Ces informations structurales sont indispensables, non seulement pour la compréhension du mécanisme de synthèse des polykétides, mais aussi pour permettre une approche de biologie synthétique efficace.

Toutes les données SAXS ont été enregistrées au synchrotron SOLEIL sur la ligne SWING qui dispose d'un système d'injection couplée à un système HPLC, ce qui permet de séparer la protéine à étudier d'éventuels agrégats, afin d'obtenir une solution monodisperse idéale pour l'enregistrement de données SAXS. Par ailleurs, les données sont enregistrées directement à plusieurs concentrations dans un pic de sortie de filtration sur gel HPLC, et il n'y a donc plus nécessité de réaliser des dilutions en série comme c'est le cas pour les analyses en batch (David, et Pérez, 2009).

Nous avons utilisé une approche de « dissection – réassemblage » pour étudier par SAXS l'ensemble du module 5. Pour cela nous avons utilisé plusieurs constructions : *KS₅-linker*, *ACP_{5a}-ACP_{5b}*, le tridomaine *KS₅-ACP_{5a}-ACP_{5b}*, et le module entier. Nous avons donc analysé les constructions d'une part pour résoudre la structure à faible résolution du module mais aussi pour avoir accès à des informations concernant les aspects dynamiques des domaines et des *linkers* flanquant et donc du module en lui-même. Afin de placer les différents domaines au sein des enveloppes SAXS, nous avons également eu recours à des méthodes permettant de disposer de

modèles à haute résolution, comme la RMN pour la résolution de la structure des ACP par, et la modélisation par homologie pour le domaine KS qui ne cristallise pas.

Nous avons tout d'abord construit un modèle pour le KS₅-*linker* par homologie de séquence avec la structure cristallographique du KS-AT du module 5 de la PKS synthétisant l'érythromycine A (PDB : 1GH5 ; Tang, *et al.*, 2006). Plusieurs logiciels de modélisations par homologie de séquence avaient permis de générer une dizaine de modèles. Ces modèles ont ensuite été discriminés en calculant leur courbe de diffusion théorique qui a été comparée à la courbe de diffusion expérimentale enregistrés pour le KS₅-*linker*. Ainsi un modèle en adéquation avec les données SAXS a pu être généré ($\chi^2= 1,58$) (Article, Fig 3). Le modèle montre que le domaine KS est bien structuré et sous la forme d'un homodimère comme attendu (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007). Les données SAXS ont également permis de montrer que le *linker* en aval du KS est bien structuré. Ces résultats sont confortés par des études menés sur des régions comparables sur d'autre systèmes PKS (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007 ; Bretschneider, *et al.*, 2013 ; Gay, *et al.*, 2016). L'analyse de la superposition des courbes de diffusion théoriques et expérimentales montrait quelques différences semblant suggérer une certaine flexibilité de la protéine. Grâce aux calculs d'enveloppe et à de la modélisation en corps rigide, nous avons mis en évidence une rotation du *linker* KS-AT par rapport au domaine KS. Cette flexibilité n'avait jamais été mise en évidence auparavant par les structures cristallographiques des KS résolues jusqu'alors (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007 ; Bretschneider, *et al.*, 2013). Cela démontre l'utilité du SAXS pour l'étude de systèmes dynamiques.

Nous avons également analysé par SAXS la construction ACP_{5a}-ACP_{5b} où les deux ACP semblent être indépendants. Pour pouvoir interpréter les données, les structures des deux ACP du module 5 (ACP_{5a} et ACP_{5b}) ont été résolues par RMN par H. Rabeharindranto et le Dr B. Chagot. En effet, les deux ACP sont séparés par un *linker* de 28 acides aminés qui adopte une conformation relativement étendue. Nous avons pu déterminer pour cette construction un D_{\max} de 103 Å, ce qui est conséquent pour une protéine de 187 résidus. Ceci confirme que les deux ACP sont distants l'un de l'autre, et n'interagissent pas ensemble. Ce résultat a été conforté par une analyse RMN du didomaine ACP_{5a}-ACP_{5b} montrant que les deux ACP sont clairement indépendants. En effet, seul un décalage dans les déplacements chimiques des résidus au voisinage du *linker* reliant les deux ACP a été constaté en comparant les

spectres des ACP_{5a} et ACP_{5b} à celui de l'ACP_{5a}-ACP_{5b}. L'analyse par SAXS nous a également permis de constater que l'un des deux ACP adopte plusieurs conformations et subit un mouvement de ressort grâce au linker qui les relie. Cela donne naissance à une conformation où les deux ACP sont distants et une conformation dans laquelle les deux ACP se rapprochent l'un de l'autre (Article, Fig 4). Nous avons ainsi démontré la flexibilité du *linker* reliant les deux ACP.

Nous avons ensuite analysé par SAXS la construction KS₅-ACP_{5a}-ACP_{5b}, c'est-à-dire le module 5 sans le domaine C-terminal de dimérisation putatif. Nous avons obtenu la structure à faible résolution d'un module arborant une conformation « bras ouvert ». Dans ce module, le domaine KS₅ occupe une position centrale, l'ACP_{5a} est adjacent au domaine KS₅, et l'ACP_{5b} est situé au bout du long *linker* flexible, et semble ainsi plus accessible aux enzymes de la cassette de β -méthylation que l'ACP_{5a}.

Enfin, nous avons étudié quelle était l'influence du domaine C-terminal de ce module, prédit comme étant un domaine de dimérisation (Gu, *et al.*, 2011). Les données SAXS ont montré que ce domaine n'est pas dimérique dans le contexte du module entier. Nous avons donc émis l'hypothèse qu'il s'agisse d'un domaine de docking (Broadhurst, *et al.*, 2003 ; Buchholz, *et al.*, 2009 ; Whicher, *et al.*, 2013) nécessaire pour l'interaction du module 5 porté par la chaîne polypeptidique VirA avec le module 6 porté par la protéine VirFG.

Dans ce travail, les séquences codant les protéines ont été clonées, et les protéines produites et purifiées par le Dr J. Davison, la partie RMN a été menée par le Dr B. Chagot et H. Rabeharindranto. J'ai pour ma part réalisé les études SAXS sous la direction du Dr. A. Gruez et du Pr. K.J. Weissman.

Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module†

Cite this: *Chem. Sci.*, 2014, 5, 3081

Jack Davison, Jonathan Dorival, Hery Rabeharindranto,‡ Hortense Mazon, Benjamin Chagot,* Arnaud Gruez* and Kira J. Weissman*

The modular polyketide synthases (PKS) are gigantic multienzymes which synthesize diverse secondary metabolites of therapeutic value. Although structural information is increasingly available for the 'cis-AT' class of modular PKS, almost nothing is known about the evolutionarily divergent 'trans-AT' PKS, which characteristically incorporate an iteratively-acting acyl transferase. We report here the SAXS solution structure of a complete *apo* module from the virginiamycin *trans*-AT PKS, which is fundamentally different to structural models proposed for the *cis*-AT PKS based on the crystal structure of animal fatty acid synthase. The module, which serves as a platform for the β -modification of the polyketide intermediate, consists of a ketosynthase (KS) and two acyl carrier protein (ACP) domains. In our solved structure, the homodimeric KS, which is flanked by well-folded linker regions, occupies the center of the module. While the first ACP is located close to the KS, the second is situated at the end of a flexible linker, and mobile. Taken together, these data provide a physical explanation for the functional non-equivalence previously observed for certain tandem ACPs of *trans*-AT PKS. Furthermore, the overall open shape of the module renders the second ACP highly accessible, which may be critical for its interaction with multiple *in trans* catalytic partners. Finally, our analysis redefines the function of a putative dimerization motif of tandem ACPs as a docking domain, suggesting that the module likely adopts a more closed form in order to affect transfer of the chain extension intermediate to the subsequent module.

Received 24th December 2013

Accepted 3rd April 2014

DOI: 10.1039/c3sc53511h

www.rsc.org/chemicalscience

Introduction

A multitude of bioactive polyketide natural products are biosynthesized using a division-of-labor strategy, in which the discrete steps in assembling the chains are carried out by individually-folding protein domains.¹ The domains are organized into functional units called 'modules' within large multienzyme polyketide synthases (PKS), where each module carries out one chain extension cycle. Thus, each module contains an acyl transferase (AT) domain to choose the appropriate building block and a ketosynthase (KS) domain to attach it to the growing chain using Claisen-type chemistry. A set of optional domains (ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)) then introduces the appropriate level of reduction at the initially-formed β -keto center. Finally, a

dedicated thioesterase (TE) at the end of the last PKS multienzyme liberates the fully-processed chain either by macrocyclization or hydrolysis. The efficacy of this assembly-line strategy is augmented by the covalent tethering of the chain-extension intermediates to non-catalytic acyl carrier protein (ACP) domains, *via* a phosphopantetheine (Ppant) prosthetic group.²

Presumably in reflection of the efficiency of this mode of biosynthesis for complex secondary metabolites, Nature has convergently evolved two distinct classes of modular PKS. In the well-studied *cis*-AT systems, such as the prototypical PKS (DEBS) responsible for erythromycin assembly (Fig. 1a),³ the ATs are present as domains within the primary sequence of each module. For such PKS, there is a strong correlation between the domain composition of the modules and the structure of the resulting chain extension intermediates. In contrast, in the more-recently discovered *trans*-AT PKS, the ATs are discrete enzymes which act iteratively to furnish a single extender unit to each module (Fig. 1b).⁴ As a result of further organizational features of the *trans*-AT pathways – the presence of duplicated and inactive domains, modules divided between two proteins, participation of large numbers of additional *trans*-acting enzymes, *etc.* – there is no strict co-linearity between the primary sequence of the PKS and the product structures.⁵

Molecular and Structural Enzymology Group, Université de Lorraine, IMoPA, UMR 7365, Vandœuvre-Lès-Nancy, F-54500, France. E-mail: benjamin.chagot@univ-lorraine.fr; arnaud.gruez@univ-lorraine.fr; kira.weissman@univ-lorraine.fr

† Electronic supplementary information (ESI) available: Nine figures, three tables, and supplementary references. See DOI: 10.1039/c3sc53511h

‡ Present address: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés – INSA, UMR INSA/CNRS 5504 – UMR INSA/INRA 792, 135 Avenue de Rangueil, 31077 Toulouse CEDEX 4, France.

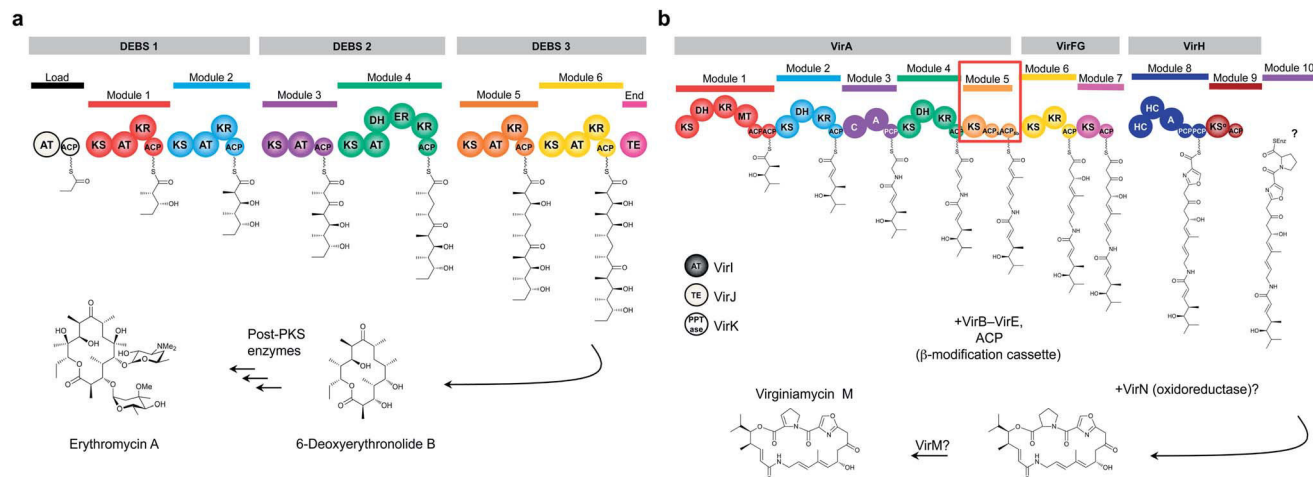


Fig. 1 Schema of the biosynthesis by representative *cis*-AT and *trans*-AT PKS systems. (a) The *cis*-AT PKS which assembles erythromycin A consists of three gigantic multienzymes (DEBS 1, DEBS 2 and DEBS 3) comprising loading, chain extension and termination modules. For this system, there is co-linearity between the order and domain composition of the modules and the structure of the resulting chain-extension intermediate. (b) The *trans*-AT PKS responsible for virginiamycin biosynthesis includes at least nine chain extension modules (seven polyketide synthase and two non-ribosomal peptide synthetase); the presumed tenth module for incorporation of proline was not identified in the sequenced region.⁸ In the course of this work, we determined that modules 6 and 7 are located on a single protein, renamed VirFG (for details see the Methods). This system exhibits several typical characteristics of the *trans*-AT PKS, including a discrete, iteratively acting AT (VirI), duplicated domains (ACPs of modules 1 and 5, PCPs of module 8), a condensation-inactive module (module 9), etc. In this study, we targeted module 5 (boxed) for structure elucidation by SAXS. Key: AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase; MT, methyltransferase; C, condensation; A, adenylation; PCP, peptidyl carrier protein; HC, condensation/heterocyclization; KS^o, inactive ketosynthase.

Another notable feature of these systems, which is shared by certain *cis*-AT PKS, is the presence of nonribosomal peptide synthetase (NRPS) modules allowing the incorporation and modification of amino acid building blocks (Fig. 1b). Despite the divergent evolutionary origins of the *trans*-AT PKS, their products exhibit remarkable structural similarity to those of the *cis*-AT PKS, and are likewise highly bioactive. Notable metabolites produced by *trans*-AT PKS include the anti-cancer bryostatins⁶ and the clinically-used antibiotics mupirocin⁷ and virginiamycin M.⁸

There is significant interest in the three-dimensional architectures of both *cis*- and *trans*-AT PKS, as such information is not only essential to understanding their complex mode of operation, but is vital to ongoing attempts to genetically re-engineer the systems to produce analogues of potential medicinal value.⁹ Resolution of the crystal or NMR structure of each of the domains of *cis*-AT PKS¹⁰ has revealed that the proteins have similar folds to the corresponding domains of animal fatty acid synthase (FAS), although there are notable differences in oligomerization state (for example, the ER of FAS is dimeric, while the PKS ER is monomeric¹¹). Taken together, these data have been used to propose several models for the overall structures of *cis*-AT PKS modules incorporating various complements of reductive activities,¹¹ by extrapolation from the 3.2 Å resolution crystal structure of porcine FAS.¹² No such model exists for *trans*-AT PKS, however, as only the structures of a *trans*-AT domain,¹³ an ACP-ACP didomain,¹⁴ and a KS-branching (B) didomain¹⁵ have been solved to date. We report here the first solution structure of a complete *apo* module from a model *trans*-AT system comprising KS and tandem ACP

domains, obtained by small-angle X-ray scattering (SAXS)¹⁶ in combination with NMR and homology modeling.

Results

Strategy for resolution of the structure of a PKS module by SAXS

No three-dimensional crystallographic structure of an intact PKS module or multi-modular subunit has been published to date, presumably due to the high, inherent flexibility of the proteins.¹² We thus aimed to use SAXS to solve the structure of a representative module from a *trans*-AT PKS. SAXS has recently come to the fore as a powerful means to analyze large, flexible multidomain proteins and macromolecular complexes in solution due to advances in data collection and analysis, which allow molecular shapes to be identified accurately.¹⁶

As a representative model system, we chose the *ca.* 100 kDa module 5 of the virginiamycin NRPS-PKS of *Streptomyces virginiae* (Fig. 1b), which comprises a KS, two consecutive ACP domains (ACP_{5a} and ACP_{5b}), and a putative C-terminal dimerization motif (DD) (see ESI, Fig. S1†),⁸ which are separated from each other by linker regions of variable length (Fig. 2). Module 5 interacts with a large number of catalytic partners both in *cis* and in *trans*: the ACP of module 4, the KS of module 6, the *trans*-AT VirI, the phosphopantetheinyl transferase VirK, the proof-reading thioesterase VirJ, and a cassette of enzymes (VirC-E) that introduces a β-methyl group into the chain-extension intermediate. Modules consisting of a KS and two tandem ACPs are present in fully a third of all *trans*-AT PKS characterized to date,^{4,17–22} and such duplicated ACPs are notably common in

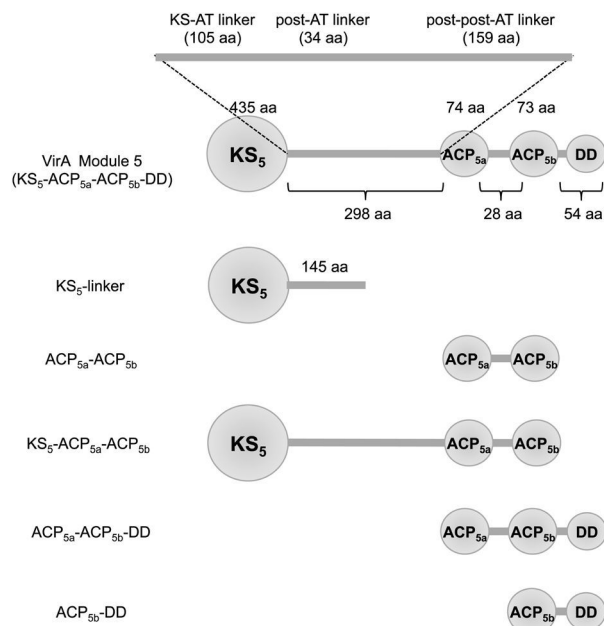


Fig. 2 Schematic of the constructs studied in this work. Six different constructs were evaluated by SAXS, including the entirety of module 5 (KS_5 - ACP_{5a} - ACP_{5b} -DD), the KS and a portion of the downstream linker (KS_5 -linker), the tandem of ACP domains and their intervening linker (ACP_{5a} - ACP_{5b}), and a construct encompassing all three domains but missing the putative dimerization region (KS_5 - ACP_{5a} - ACP_{5b}). In order to further define the role of the dimerization domain, we also analyzed by gel filtration and SAXS, the tandem of ACPs and ACP_{5b} in the presence of the dimerization region (ACP_{5a} - ACP_{5b} -DD and ACP_{5b} -DD, respectively). All ACP-containing constructs were studied in their *apo* forms. The number of residues comprising the various domains and linkers is indicated.

modules of *trans*-AT PKS in which the intermediate undergoes a β -methylation.²³ Our results are also likely to be relevant to simpler KS-ACP modules (see for example module 7 of VirFG (Fig. 1b)).⁴

In order to reconstitute the structure of the entire module from SAXS data,¹⁶ we used a ‘dissect and build’ strategy, using SAXS to study first KS_5 and a portion of the downstream linker (KS_5 -linker), then the tandem ACPs (ACP_{5a} - ACP_{5b}), the tridomain KS_5 - ACP_{5a} - ACP_{5b} , and finally the complete module (KS_5 - ACP_{5a} - ACP_{5b} -DD) (Fig. 2). Each of the proteins was expressed in its *apo* form in *E. coli* and purified to homogeneity (Fig. S2†). Attempts to obtain KS_5 - ACP_{5a} were unsuccessful, due to poor solubility (data not shown).

Structural analysis of the KS_5 -linker by SAXS

Modules 4 and 5 are joined within VirA by an approximately 55 residue linker (GenBank: BAF50727.1). The N-terminal boundary of KS_5 was chosen to lie within this linker, just upstream of the conserved domain as identified in the crystal structures of the KS-AT didomains from DEBS modules 3 and 5.^{24,25} The region downstream of the KS, which is conserved among *trans*-AT PKS,²⁶ has previously been described as a non-functional remnant of an AT domain, and suggested to play a role in docking the *trans*-AT with the PKS. Our reanalysis of the

298 amino acids separating the KS from ACP_{5a} revealed consecutive regions within the first 139 amino acids with convincing homology to the so-called KS-AT and post-AT linkers identified in the DEBS KS-AT didomain structures^{24,25} (Fig. S3†). In these didomains, the KS-AT linker which adopts an $\alpha\beta$ fold, and the post-AT linker which wraps back over both the AT domain and the KS-AT linker to interact with the KS, serve to fix the relative positions of the KS and AT domains. We therefore hypothesized that the KS and these 139 amino acids might form a structural unit. Consistent with this idea, the resulting construct which spanned these regions was obtained in satisfactory yield (6 mg L^{-1}), and shown to be homodimeric as expected²⁷ by native mass spectrometry (calculated dimer MW = 129 461.2 Da; found = 129 480 Da; monomer MW by denaturing mass spectrometry = 64 730.0 Da (calcd = 64 730.6 Da)).

The SAXS data on the KS_5 -linker were collected on the SWING beamline at the Synchrotron SOLEIL, which allows for online sample purification by gel filtration just prior to analysis. In this way, it is possible to select a population of molecules within the sample which have the same gyration radius (R_g), or average protein dimension. The buffer conditions used in our experiments are known to promote the catalytic activity of both *cis*- and *trans*-AT PKS.^{28–30} Guinier’s plot of the SAXS data acquired on the most abundant conformer of the KS_5 -linker gave an R_g of $33.1 \pm 0.5 \text{ \AA}$, while calculation of the distance distribution function ($P(r)$) using GNOM³¹ yielded a maximum particle diameter (D_{max}) of 105 \AA (Fig. 3a and b; all SAXS data are summarized in Table S1†). These data and a molecular weight of 137.2 kDa calculated using SAXSMOW³² are consistent with a globular, homodimeric protein (monomer MW = 64.8 kDa).

We then derived *ab initio* three-dimensional envelopes from the experimental data using the program DAMMIN³¹ with imposition of second-order symmetry (Fig. 4a). In order to place the KS domain within the average envelope using SUPCOMB³³ in the absence of its high-resolution structure, we generated a homology model of the domain and adjacent linkers (Fig. 3c) with MODELLER³⁴ using KS_5 of DEBS as a template (40% sequence identity). The resulting model was then dimerized by alignment of the monomers with the structure of the DEBS KS_5 -AT₅ didomain (PDB: 2HG4²⁴). (During preparation of this manuscript, the crystal structure of the didomain KS-branching (B) enzyme from the RhiE subunit of the rhizoxin *trans*-AT PKS was published.¹⁵ However, this KS was not a better model for the VirA KS_5 , and so all analysis was based on the VirA KS_5 -linker model derived from DEBS as template (for further discussion on this point, see the Methods)).

For values above 0.125 \AA^{-1} (Fig. 4a), there is a minor divergence between the calculated and experimental curves, which suggests that a portion of the KS_5 -linker construct adopts multiple orientations in solution. Specifically, the different *ab initio* shape calculations reveal small differences in the molecular envelope corresponding to the linker region, suggesting that it is this part of the structure which is flexible. To validate this idea, we aimed to determine whether changing the position of the KS-AT linker relative to KS_5 would improve the agreement between the calculated and experimental data.

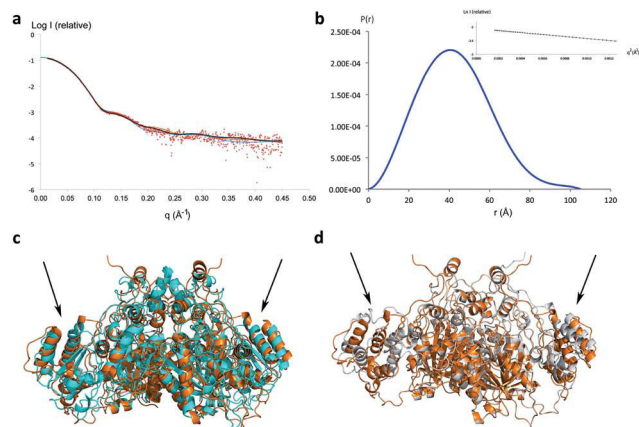


Fig. 3 SAXS analysis of the KS₅-linker. (a) Superimposition of the experimental SAXS data (red dots) with the theoretical curves for the KS₅-linker, calculated on the basis of the solved structures of the DEBS KS₅ (orange) and RhiE KS (cyan) domains as templates for the homology model. The curve in black corresponds to the theoretical curve for the KS₅-linker model derived from DEBS KS₅ as template, after varying the relative orientations of the KS and the KS-AT linker with SASREF. Improved agreement was obtained by rotating the two KS-AT linkers relative to their original orientations by approximately 18 and 21°, respectively, such that they occupy non-equivalent positions relative to the KS dimer. (b) The distance distribution function derived for the KS₅-linker calculated with GNOM. Inset is the Guinier plot. (c) Superimposition of the homology models of the KS₅-linker (orange) generated using DEBS KS₅ (orange) and RhiE KS (cyan) as template. (d) Superimposition of the DEBS-derived homology model of the KS₅-linker (orange) with that obtained from SASREF (grey), illustrating the change in relative position of the KS-AT linkers (compare the regions indicated with the arrows in panels c and d).

To this end, we performed with SASREF³⁵ a rigid body simulated annealing using a dimeric KS and the two linkers as search models. Improved agreement was obtained for a model in which the two KS-AT linkers are rotated relative to their original orientations by approximately 18 and 21°, respectively, and occupy apparently non-equivalent positions relative to the KS dimer (compare Fig. 3c and d). In contrast, comparison of the crystal structures of the KS-AT linkers from DEBS²⁴ and RhiE¹⁵ shows that their positions superimpose. Taken together, these data underline the power of SAXS relative to crystallography to reveal protein conformational flexibility in solution.

Structural analysis of ACP_{5a}-ACP_{5b} by SAXS and NMR

The tandem ACPs 5a and 5b, which exhibit 52% mutual sequence identity, are separated by a region of 28 residues rich in Ala and Pro (68%), characteristic of extended, but flexible linkers³⁶ (Fig. S4a†). SAXS analysis of a construct encompassing the two ACPs in their *apo* forms yielded an R_g of 29.0 ± 0.2 Å, and a D_{max} of 103 Å, which are large for a 187-amino acid protein. The distance distribution function (Fig. S5a†) revealed two peaks, consistent with a dumbbell-like architecture, in which two independent domains are joined by an unstructured region. Independent confirmation of this overall shape was provided by the recent publication of the NMR structure of an

apo ACP-ACP didomain from subunit MmpA of the mupirocin *trans*-AT PKS, which also incorporates a 28-residue ACP-ACP linker (Fig. S4a†).¹⁴

We also solved the NMR solution structures of the two *apo* ACP domains (Fig. 5 and Tables S2 and S3†). The structures of the ACP_{5a} and ACP_{5b} domains exhibit a typical ACP fold composed of three major antiparallel α helices (I, II and IV) and a short perpendicular α helix (III). The helices in ACP_{5a} and ACP_{5b} are as follows: helix I, R6734-V6749/A6837-T6850; helix II, S6770-A6783/S6871-Y6885; helix III, A6792-D6795/I6893-D6896; and helix IV P6799-L6810/L6900-R6910. The long loops between helix I and II of ACP_{5a} and ACP_{5b} include two short helical regions (P6754-R6756 and F6763-L6766 for ACP_{5a} and P6855-D6857 and F6864-D6857 for ACP_{5b}). The Ser which serves as the point of attachment of the phosphopantetheine prosthetic group (Ser6770 and S6871, respectively) is located at the N-terminus of helix II. These two structures allowed the sizes of the ACPs to be estimated at 35 Å, which when taken together with the SAXS results, reveal that the intervening linker is approximately 33 Å in length. Analysis of 19 additional sequences separating ACP domains in *trans*-AT PKS (Fig. S4a†) shows that these regions vary in length from 12 to 72 residues, and exhibit no strong homology. However, in common with the Vir ACP_{5a}-ACP_{5b} linker, many of the sequences, including that of MmpA, are biased towards Ala and Pro, suggesting that the global properties of the linkers may be shared. Indeed, the linker regions connecting the five tandem ACPs of the polyunsaturated fatty acid (PUFA) synthase of *Photobacterium profundum* also show this biased amino acid composition (Fig. S4b†), and the overall protein adopts an elongated, 'beads-on-a-string' architecture.³⁷

Using SUPCOMB, we next fitted the NMR structures of the two *apo* domains into the average of ten *ab initio* molecular envelopes ($\chi^2 = 1.785$) (Fig. 4b). This analysis showed that the position of one of the ACPs is well-defined, while that of the second is not. Indeed, the shape of the average molecular envelope reveals a spring-like movement along the inter-ACP axis, and thus that the intervening linker is flexible. Structural superimposition of one ACP on the didomain using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) confirms the locational variability of the second ACP, which occupies several positions relative to the other, giving rise to overall compact or extended shapes. The relative percentages of these two configurations were estimated at 80% and 20%, respectively, using OLIGOMER (<http://www.embl-hamburg.de/biosaxs/manuals/oligomer.html>). Notably, this finding is consistent with the form of the $P(r)$ function, in which the second peak presents two maxima (Fig. S5a†).

Our SAXS results additionally suggest that the ACPs do not interact at least in their *apo* forms. In order to confirm this result, we analyzed the *apo* ACP_{5a}-ACP_{5b} didomain by NMR. Comparison of the [¹H, ¹⁵N]-HSQC spectrum of *apo* ACP_{5a}-ACP_{5b} with those of the individual *apo* ACPs revealed no significant differences in chemical shift for key residues, with the exception of amino acids located at or near the covalent linkage between the two domains (Fig. 5b). These data are

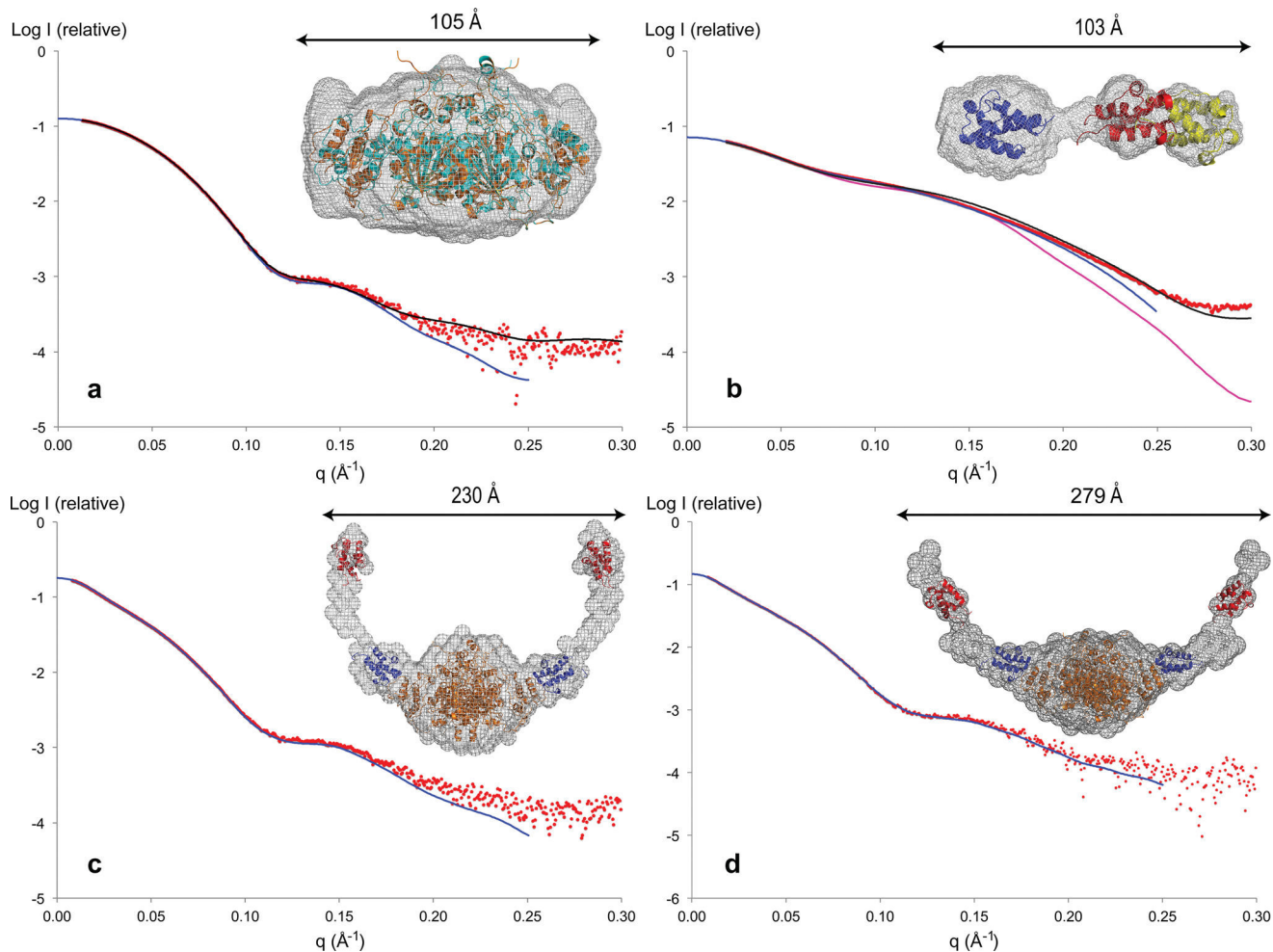


Fig. 4 SAXS profiles and domain organization within module 5. (a) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data obtained on the KS_5 -linker (red dots). The fit to the data of the KS_5 -linker model generated by rigid body modeling using SASREF is represented in black. Inset, the *ab initio* envelope of the KS_5 -linker construct calculated from DAMMIN (grey mesh) is shown in superimposition by SUPCOMB with the KS_5 -linker model obtained using the DEBS KS_5 -linker as template (orange), and the model calculated from SASREF (cyan). (b) Fit between the *ab initio* model computed with DAMMIN (solid blue line) with the experimental SAXS data for *apo* ACP_{5a} - ACP_{5b} (red dots). Fit to the data of the *apo* ACP_{5a} - ACP_{5b} model in its compact conformation (pink) calculated by rigid body modeling using SASREF. Fit to the data of the theoretical curve computed with OLIGOMER of an 80 : 20 mixture of the compact and extended forms of the *apo* ACP_{5a} - ACP_{5b} di-domain (black). Inset, the *ab initio* envelope of *apo* ACP_{5a} - ACP_{5b} calculated from DAMMIN (grey mesh) with superimposition of one *apo* ACP (blue) and the second *apo* ACP in two locations (red and yellow) using SUPCOMB. (c) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data acquired on the *apo* KS_5 - ACP_{5a} - ACP_{5b} (red dots). Inset, the *ab initio* envelope of the *apo* KS_5 - ACP_{5a} - ACP_{5b} construct calculated from DAMMIN (grey mesh) is shown with superimposition by SUPCOMB of the KS_5 -linker (orange), *apo* ACP_{5a} (blue) and *apo* ACP_{5b} (red). (d) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data obtained on *apo* KS_5 - ACP_{5a} - ACP_{5b} -DD (red dots). Inset, the *ab initio* envelope of the *apo* KS_5 - ACP_{5a} - ACP_{5b} -DD construct calculated using DAMMIN (grey mesh) with superimposition of the KS_5 -linker (orange), *apo* ACP_{5a} (blue) and *apo* ACP_{5b} (red).

consistent with the absence of any direct contact between the tethered ACPs, as found for the ACPs of MmpA¹⁴ and the PUFA synthase.³⁷

Structural analysis of KS_5 - ACP_{5a} - ACP_{5b} by SAXS

We next analyzed by SAXS an *apo* construct incorporating all of the functional domains of module 5 in the absence of the putative C-terminal dimerization region, giving an R_g of 56.5 ± 0.7 Å and a D_{max} of 230 Å (Fig. S5b†). An average *ab initio* three-dimensional envelope was then calculated from the experimental data with DAMMIN, with imposition of second-

order symmetry (Fig. 4c). This parameter is justified by the homodimeric nature of the KS and by the fact that lack of such symmetry imposition caused the calculation to diverge, likely due to the large number of amino acids in the protein. Use of MONSA,³⁵ then allowed us to fit simultaneously the obtained diffusion data on the KS_5 -linker and the *apo* KS_5 - ACP_{5a} - ACP_{5b} constructs into the molecular shape, clearly identifying the position of the dimeric KS domain at the center of the protein ($\chi^2 = 1.885$). The two *apo* ACP NMR structures were then placed into the envelope using as a constraint the D_{max} value of 103 Å obtained for the *apo* ACP_{5a} - ACP_{5b} didomain.

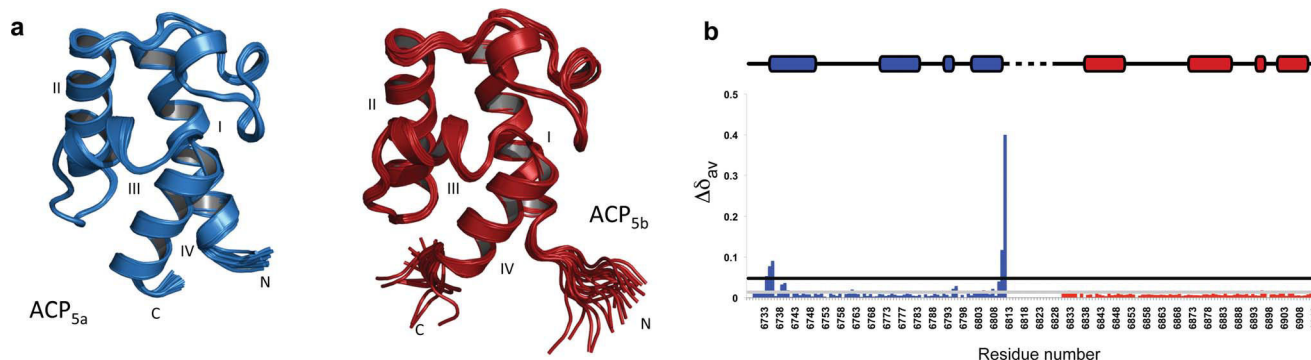


Fig. 5 NMR analysis of *apo* ACP_{5a}, *apo* ACP_{5b} and *apo* ACP_{5a}-ACP_{5b}. (a) The three-dimensional solution structures of the *apo* ACP_{5a} and ACP_{5b} domains were determined using multi-dimensional heteronuclear NMR spectroscopy. Structure calculation was carried out using a two-step procedure consisting of initial structure generation using CYANA followed by restrained molecular dynamics refinement within Amber. The quality of the structures was assessed using PROCHECK-NMR and the molprobiy server. Shown are the final ensembles of 20 conformers representing the solution structures of *apo* ACP_{5a} (blue) and *apo* ACP_{5b} (red), depicted with all heavy atoms. As expected based on their sequence identity (52%), *apo* ACP_{5a} and ACP_{5b} have similar structures (RMSD over 72 residues is 1.2 Å). (b) Average ¹⁵N-¹H chemical shift changes ($\Delta\delta_{av}$) plotted as a function of ACP_{5a}-ACP_{5b} residue number (sequence of VirA). The secondary structure elements present in *apo* ACP_{5a} and ACP_{5b} are shown above the plot (the linker is indicated by a dotted line as its secondary structure is unknown). ¹⁵N-¹H chemical shift differences are calculated between chemical shift of equivalent residues in *apo* ACP_{5a}-ACP_{5b} and ACP_{5a} (blue) and between the chemical shift of equivalent residues in *apo* ACP_{5a}-ACP_{5b} and ACP_{5b} (red). The grey and black lines indicate the mean average chemical shift and the average plus one standard deviation, respectively, and were used to identify significant changes. This analysis demonstrates that the chemical shifts of only a few residues vary between the isolated *apo* ACP domains and the didomain construct.

This analysis revealed that, together, the KS and the two ACP domains form an ‘open arms’ structure, wherein ACP_{5a} is located directly adjacent to the KS, and ACP_{5b} lies at the extremity of the arm. The respective locations of the two ACPs suggests that the position of ACP_{5a} is relatively fixed, whereas that of ACP_{5b} is much less well-defined, consistent with the inherent mobility observed for one of the two ACPs of ACP_{5a}-ACP_{5b}. Nonetheless, as ACP_{5a} sits at a distance of approximately 50 Å from the KS (as measured from the KS active site Cys to the Ser of the ACP which is modified by phosphopantetheine), which is out of reach of the 20 Å long prosthetic group, an interaction between the two domains would necessitate a structural rearrangement.

The other notable result from the form calculation is that the 159-residue region C-terminal of the KS-AT and post-AT linkers (designated as the ‘post-post-AT linker’ (Fig. 2)) must adopt a compact structure. It was not possible to model this region, however, due to the complete absence of homologous sequences in the protein database. Nonetheless, inspection of the linker reveals a high proportion of residues which are correlated with natively unfolded proteins (*i.e.* Q, S, P, E, K, G and A;³⁸ 102/159 residues), an analysis supported by disorder prediction using PONDR-FIT³⁹ (Fig. S6a and b†). Such regions often adopt a defined fold only in the presence of other proteins, and are implicated in molecular recognition processes with multiple partners.⁴⁰ To evaluate the generality of this finding, we analyzed the comparable region from 35 additional *trans*-AT PKS modules containing consecutive KS and ACP domains (Fig. S6c†). Although no clear homology to the VirA linker was detected in these sequences (with the exception of the corresponding linker from the pristinaamycin II cluster of *Streptomyces pristinaespiralis* which generates the same

metabolite⁴¹ (Fig. S6d†)) nor any mutual sequence similarity, a large proportion of the regions also are predicted by PONDR-FIT to adopt an unfolded state (Fig. S6e†).

Analysis of KS₅-ACP_{5a}-ACP_{5b}-DD

In silico analysis of module 5 revealed a 50-residue C-terminal sequence with homology to a 51-amino acid region from the CurA subunit of the curacin *cis*-AT PKS (Fig. S1†). Notably, CurA contains a triplet of ACP domains and is also the site of β -modification of the polyketide intermediate. As this region was shown to increase the proportion of dimeric relative to monomeric forms of constructs incorporating the three tandem ACPs,⁴² it was designated as a dimerization motif. To evaluate the effect of this putative dimerization element on the overall structure of module 5, we compared the SAXS data obtained on the entire *apo* module (KS₅-ACP_{5a}-ACP_{5b}-DD) to that of the *apo* KS₅-ACP_{5a}-ACP_{5b} construct. Guinier’s plot yielded an R_g of 69 ± 1 Å, while the distance distribution function produced a D_{max} of 279 Å (Fig. S5c†). As both parameters are higher than that of KS₅-ACP_{5a}-ACP_{5b}, the SAXS analysis revealed that, contrary to expectation, the major conformer of the *apo* module also adopts an open structure in the presence of the putative dimerization domain. An average *ab initio* three-dimensional envelope was then calculated from the experimental data with DAMMIN ($\chi^2 = 2.204$), with imposition of second-order symmetry (Fig. 4d). The structures of the KS₅-linker and the two *apo* ACPs were then placed within the obtained form by comparison to the data obtained on *apo* KS₅-ACP_{5a}-ACP_{5b}. The important result of this analysis is that the putative dimerization domains on each polypeptide do not interact with each other and that, indeed, the module still assumes an open shape. Furthermore,

comparison of the structures of *apo* KS₅-ACP_{5a}-ACP_{5b} and *apo* KS₅-ACP_{5a}-ACP_{5b}-DD shows that the 'arms' can move, leading to relatively more open or closed conformations.

This observed lack of dimerization prompted us to determine the oligomerization state of further constructs incorporating the C-terminal dimerization motif and the tandem ACP domains (ACP_{5a}-ACP_{5b}-DD) or the second ACP alone (ACP_{5b}-DD). Both constructs were analyzed in their *apo* forms. Analysis by analytical gel filtration demonstrated that while *apo* ACP_{5a}-ACP_{5b} is a monomer, both *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD elute as oligomers (Fig. S7a-d†). Independent analysis, however, by native mass spectrometry (Fig. S7e-h†) showed that both constructs incorporating the DD domain maintained the largely monomeric structure of *apo* ACP_{5a}-ACP_{5b}. To understand this discrepancy, we analyzed *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD by SAXS (Fig. 6). This experiment yielded an R_g of 37.0 ± 0.3 Å and a D_{max} of 140 Å for ACP_{5b}-DD (relative to $R_g = 11.1$ Å and $D_{max} = 36.4$ Å for *apo* ACP_{5b} alone, calculated with CRY SOL⁴³ based on the NMR structure), and an R_g of 43.5 ± 0.2 Å and a D_{max} of 170 Å for ACP_{5a}-ACP_{5b}-DD (Fig. S5d-g†).

We calculated the *ab initio* molecular forms of the two proteins using DAMMIN with and without imposition of

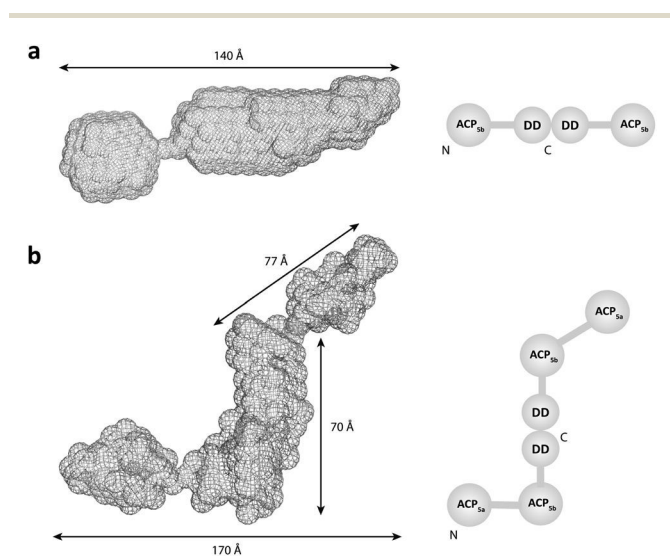


Fig. 6 Molecular envelopes of *apo* ACP_{5b}-DD and ACP_{5a}-ACP_{5b}-DD. (a) The *ab initio* envelope of the *apo* ACP_{5b}-DD calculated with DAMMIN in the absence of imposition of symmetry, shows that the construct adopts an extended, largely cylindrical shape, in which the ACPs are located at the extremities, and the two docking domains dimerize at the center. All of the calculations used to derive the average envelope show that the ACPs are not symmetrical, as the position of one of the two is less well-defined. (b) The *ab initio* envelope of the *apo* ACP_{5a}-ACP_{5b}-DD construct derived with DAMMIN in the absence of imposed symmetry. Measured dimensions of the shapes are indicated with arrows. Interestingly, the SAXS analysis shows that while the ACP_{5b}-DD portion of ACP_{5a}-ACP_{5b}-DD assumes a conformation similar to that in the structure of ACP_{5b}-DD alone, the ACP_{5b}-DD/DD-ACP_{5b} distance is reduced (70 vs. 140 Å). Similarly, the ACP_{5a}-ACP_{5b} distance (77 Å) shows that this region of the structure is also compacted relative to ACP_{5a}-ACP_{5b} alone (103 Å) (Fig. 5). Thus, our SAXS data reveal that the presence of nearby partner domains and linker regions can have a significant effect on the shape and flexibility of specific portions of the PKS.

second-order symmetry; the calculation performed in the absence of symmetry yielded a significantly better fit to the experimental data. We next attempted using MONSA to determine the contribution of each *apo* ACP to the three-dimensional forms of both constructs, but despite repeated rounds of calculation, this exercise was unsuccessful (for analysis of this calculation, see the Methods). Nonetheless, from the shapes, dimensions and molecular weights of the molecules calculated using SAXSMOW³² (*apo* ACP_{5b}-DD: calcd MW = 14.3 kDa, found = 31.4 kDa; *apo* ACP_{5a}-ACP_{5b}-DD: calcd = 24.5 kDa, found = 48.9 kDa) it is clear that both proteins are dimeric, and adopt extended structures *via* dimerization of the DD regions (Fig. 6). In view of this result, we interpret the failure to detect dimers by mass spectrometry as evidence that the interfaces have significant hydrophobic character, as such contacts are often lost under the gas-phase conditions of the analysis.⁴⁴

Although the DD was shown to dimerize small, unconstrained proteins such as ACP_{5b}, it had no effect on the much larger KS₅-ACP_{5a}-ACP_{5b} tridomain, leading to the conclusion that the mechanism of dimerization is conformationally unavailable in the native system. This non-native mode of dimerization prompted us to consider whether the DD region might instead play a role in docking the VirA protein with its downstream partner, VirFG (Fig. 1b). Analysis of the N-terminus of VirFG identifies 70 residues upstream of the conserved KS domain (Fig. S8†). While the end of this region shows sequence characteristics of an intrinsically disordered protein (see the Methods for the sequence of VirFG), the first 28 residues exhibit convincing sequence homology to the VirA DD region (Fig. S1a†). Thus, the non-native self-association observed for the VirA DD in the *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD constructs is likely a result of the inherent similarity between the DD and its natural VirFG N-terminal docking domain partner. Taken together, these observations are consistent with the idea that the C-terminal DD participates in a docking interaction with the N-terminus of VirFG, possibly through the formation of a four α -helical bundle – a structure which is consistent with the compact, roughly spherical shape of the DD docking complex observed by SAXS (Fig. 6).

In support of this hypothesis, the VirA DD region and the N-terminus of VirFG also show low, but convincing homology to the so-called 'class 2' C-terminal and N-terminal docking domain regions, respectively, identified in PKS subunits from a range of *cis*-AT systems in both cyanobacteria and myxobacteria^{45,46} (Fig. S1a†). Indeed, the CurA DD also shows a comparable level of sequence similarity to the C-terminal docking domains, and so should be included in this classification. The crystal structures of two such pairs of docking domains from the curacin system were recently reported,⁴⁶ revealing three important findings with relevance to our study: firstly, that the C-terminal docking domains are monomeric and so do not contribute to subunit dimerization; secondly, that docking is mediated by formation of an α -helix bundle; and finally, that the docking interface consists of a complementary hydrophobic surface, surrounded by electrostatic interactions that mediate

specificity. Taken together with our data, these results strongly suggest that *trans*-AT PKS systems also use such 'class 2'-like docking domains to ensure faithful transfer of intermediates between subunits.

Discussion

For modular PKS systems, an X-ray crystal structure of a complete module or multi-modular subunit is still awaited. In any event, as work on the vertebrate fatty acid synthase has shown,⁴⁷ static crystal structures do not capture key features of the highly dynamic native multienzyme. We demonstrate here that SAXS, although an inherently lower resolution technique, can provide invaluable information on whole-module architecture and dynamics in solution. We used a 'dissect and build' approach to solve the SAXS structure of *apo* module 5 from the virginiamycin *trans*-AT PKS; starting from the detailed study of smaller protein fragments by SAXS and allied computational methods, it proved possible to progress to analyze both the architecture and the dynamic behavior of the entire module.

Previous structural insights into modular PKS have drawn heavily on data gleaned from mono- and didomain proteins.¹⁰ In particular it has been shown that the KS domains of KS-AT didomains adopt their native homodimeric structures,^{24,25} so we anticipated that our model tridomain protein *apo* KS₅-ACP_{5a}-ACP_{5b}-DD would also reflect the native KS structure. Module 5 is located at the end of subunit VirA where it engages in chain transfer with module 6 located in the downstream subunit VirFG (Fig. 1b). This interaction likely necessitates the formation of a specific inter-subunit protein-protein complex. As we have studied isolated module 5, it is relevant to consider whether the complex with module 6 is transient (in which case a free VirA C-terminus is physiologically relevant) or long-lived. Measured K_d values for interactions between several *cis*-AT PKS subunits are in the range of 2.6–23.7 μ M,^{46,48,49} placing them in the class of transient protein-protein interactions which characteristically exhibit a mixture of oligomeric states.^{50,51} Indeed, the three subunits of the erythromycin (DEBS) PKS do not form a stable complex, at least *in vitro*.⁵² The advantage of such temporary associations is that they would allow for the exchange of damaged subunits which would otherwise block the entire assembly line. In view of measured intracellular protein concentrations (most notably 0.05–0.1 μ M for the PksR subunit of the bacillaene *trans*-AT system,⁵³ and more generally, 1 nM–10 μ M average concentration for *E. coli* proteins⁵⁴), this weak binding argues that a significant fraction of VirA module 5 will be uncomplexed at equilibrium. While the bacillaene NRPS-PKS was found to form a membrane-associated mega-complex in *Bacillus subtilis* cells,⁵³ the molecular basis for this macro-association and thus its consequences for the affinity and lifetimes of individual intersubunit interfaces in this and other actinomycete NRPS-PKS are unknown. Therefore, taken together, the available data support the validity of isolated VirA module 5 as a model system for studying the structure of a PKS module.

Our analysis of *apo* module 5 has revealed the average locations within the module of the dimeric KS and the two ACP

domains. The divergent positioning of the two ACPs raises questions as to their respective interactions with the KS, and indeed with other enzymatic partners of the module. The role of such tandem ACPs in the mupirocin *trans*-AT PKS has previously been investigated by gene inactivation experiments *in vivo*.⁵⁵ For the doublet ACP of this system, the data suggest that the two domains play essentially equivalent roles, while for the triplet cluster, the first ACP appears to be arranged in series relative to the following two. The non-equivalent locations of ACP_{5a} and ACP_{5b} relative to the KS in Vir module 5, as well as differences in their inherent mobility, suggest a physical explanation for in-series operation. In this mode of biosynthesis, ACP_{5a} would participate in chain extension with the KS, while ACP_{5b} would serve as the site for modification reactions on the extended unit, with direct transfer of the intermediate between the two ACPs.^{55,56} Nonetheless, the validity of this mechanism for the virginiamycin system remains to be directly demonstrated by analogous mutagenesis experiments *in vivo* or *in vitro*.

Our SAXS experiments have also provided important insights into the structure and function of interdomain linkers within *trans*-AT PKS. As in *cis*-AT systems, the KS domain and the first portion of its downstream linker, the sequence of which is conserved among many *trans*-AT systems (Fig. S3†),²⁶ form a structural unit. As this linker additionally shares an interface with the AT domain in *cis*-AT PKS,^{24,25} it may serve as a docking point for the *trans*-acting AT, as previously proposed.²⁶ Thus, despite the distinct evolutionary origin of the *cis*- and *trans*-AT PKS,⁵ these key operational aspects appear to be common to the two systems. The short linker joining the two ACP domains is relatively extended and flexible, allowing a significant mobility of ACP_{5b}. In contrast, the remaining 159 residues of the KS-ACP linker, which exhibit convincing sequence characteristics of a natively-disordered region,³⁸ adopt a compact structure, presumably due to contacts with other portions of the module. Although the precise lengths and sequences of these regions are not conserved among KS-ACP and KS-ACP-ACP modules in *trans*-AT PKS (Fig. S6†), their global properties appear to be shared. Taken together, these data have clear implications for attempts to reconfigure these systems by genetic engineering within the linker regions, as experiments which fail to respect the overall chemical character and function of these sequences are unlikely to succeed.

Module 5 ends in a 50-residue region which is predicted to include two terminal α -helices (Fig. S1†). The presumed helical regions within this sequence show homology to a 51-residue motif which follows a triplet of ACP domains at the end of the curacin PKS subunit CurA.⁴² As judged by analytical ultracentrifugation, CurA *apo* ACP-ACP-ACP constructs appeared to be largely monomeric whether or not this region was present. However, analysis by size-exclusion chromatography and negative-stain electron microscopy revealed that its presence led to a substantial shift in the monomer-dimer equilibrium towards dimer. Given these data, our finding that inclusion of the putative dimerization domain did not lead to closure of the module was surprising. An explanation is provided by our SAXS analysis (Fig. 6), which revealed that two

N-terminally truncated versions of the module incorporating the DD region dimerize in a non-native fashion, a mode of association which is not possible in the context of the complete module. We propose that this unnatural dimerization mimics the natural interaction between the C-terminal DD and the N-terminus of module 6, as both of these regions show homology to class 2 docking domains of *cis*-AT PKS⁴⁶ (Fig. S1†).

Modules in which β -processing reactions occur are positioned at the termini of *trans*-AT PKS subunits as in the case of the virginiamycin system, but they also occur at internal locations within the multienzymes (such as within the bacillaene, myxovirescin, pederin, onnamide, etnangien, oocydin, kalimanticin, psymberin, coralopyronin and bongkrekic acid PKSs^{4,18–21,57}). In the latter cases, the ACPs are covalently attached to the downstream KS domain of the next module. However, in these systems, the linkers joining the ACPs of β -modification modules to the downstream KS domains are on average significantly longer than the corresponding linkers following non-modifying modules (*ca.* 63 *vs.* 46 residues; Fig. S9a†). It is also striking that all of the ACP–KS linkers from *trans*-AT PKS are substantially longer than the comparable regions from *cis*-AT PKS (average length 20 amino acids, Fig. S9b†). Given that the average size of an amino acid is 3.4 Å in a maximally-extended polypeptide,⁵⁷ these ACP–KS linkers would appear to be compatible with an open conformation of the preceding β -modification modules, even if a portion of the linker adopts secondary structure. These data further support the physiological relevance of the module 5 structure observed by SAXS.

Nevertheless, it remains likely that module 5 clamps down transiently during transfer of the fully-processed intermediate to the homodimeric KS of module 6, a structural change which may be induced by formation of a compact complex between the putative C- and N-terminal docking domains.⁴⁶ Thus, overall, our SAXS data lead us to propose that the module adopts several distinct conformations from open to relatively closed. Indeed, it is possible that phosphopantetheinylation of the two ACPs to generate the *holo* module might in itself induce an extent of modular closure. On the other hand, the observed open form, in which in particular ACP_{5b} is highly accessible, may be necessary to allow one or both of the carrier protein domains to interact with their many *in trans* catalytic partners, possibly through formation of a large, multi-protein complex. Our results underscore the inherent flexibility of the ACP–ACP region, which may facilitate this suite of interactions. Full enzymatic processing of the intermediate may then trigger structural or compositional changes in the module which permit formation of a docking interaction with the downstream subunit VirFG, and efficient transfer of the chain to the module 6 KS domain. Such a mechanism, when coupled with the intrinsic substrate specificity of the receiving KS,⁵ could aid in programming the biosynthesis, ensuring that the full set of modification reactions occurs prior to transacylation. Experiments to evaluate this hypothesis are under way. In the meantime, it is worth emphasizing that the structure of *apo* Vir module 5 differs considerably from models

proposed for the *cis*-AT PKS systems (compare Fig. 7a and b) in which chain assembly occurs in two discrete reaction chambers,^{11,59} underlining the necessity to study the two types of systems independently.

Methods

Cloning, expression and purification

Eight constructs (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, ACP_{5a}-ACP_{5b}-DD, KS₅-linker, KS₅-ACP_{5a}-ACP_{5b}, and KS₅-ACP_{5a}-ACP_{5b}-DD) were amplified from *S. virginiae* genomic DNA using forward (f) and reverse (r) primers incorporating *Bam*HI and *Hind*III restriction sites, respectively, and were ligated into the equivalent sites of vector pBG-106 (KS₅-linker, KS₅-ACP_{5a}-ACP_{5b}, and KS₅-ACP_{5a}-ACP_{5b}-DD) or pBG-102 (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, and ACP_{5a}-ACP_{5b}-DD). The primer sequences were as follows (the introduced restriction sites are underlined, while the bases annealing to gDNA are capitalized): KS f, (*Bam*HI) 5'-tttccgggatccGCGATCGCCGTGATCGGCGTTG-3' (used for KS₅-linker, KS₅-ACP_{5a}-ACP_{5b}, KS₅-ACP_{5a}-ACP_{5b}-DD); KS r (*Hind*III) 5'-tttccgaagctttcCAGGTGGGCGATCTGCAGGTCTGACCAG-3'; ACP_{5a} f (*Bam*HI) 5'-tttccgggatccGCGGGCCGTCAGGAGGAG-3' (used for ACP_{5a}-ACP_{5b}, ACP_{5a}-ACP_{5b}-DD); ACP_{5a} r (*Hind*III) 5'-tttccgaagctttcAGGCTCCGAGCGACTCGGC-3' (used for ACP_{5a}); ACP_{5b} f (*Bam*HI) 5'-tttccgggatccGCCGTGGCCGTCGACCCCG-3' (used for ACP_{5b}, ACP_{5b}-DD); ACP_{5b} r (*Hind*III) 5'-tttccgaagctttcGACCGGGGCGGCGCGTCC-3' (used for ACP_{5b}, ACP_{5a}-ACP_{5b}); DD r (*Hind*III) 5'-tttccgaagcttTCAGCGCCGCGGGGTGAGCAG-3' (used for ACP_{5b}-DD, ACP_{5a}-ACP_{5b}-DD, KS₅-ACP_{5a}-ACP_{5b}-DD). Vector pBG-102 codes for a SUMO-His₆ tag, while vector pBG106 codes for a His₁₀ tag and a flexible linker containing the H3C precision protease cleavage site (Dr L. Mizoue, Center for Structural Biology, Vanderbilt University). Following cleavage of the SUMO-His₆ tag, the proteins incorporate a non-native N-terminal Gly–Pro–Gly–Ser sequence. The vectors were used to transform *E. coli* BL21, and constructs were expressed as His₁₀ tagged (KS₅, KS₅-ACP_{5a}-ACP_{5b}, and KS₅-ACP_{5a}-ACP_{5b}-DD) or SUMO-His₆-tagged (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, and ACP_{5a}-ACP_{5b}-DD) fusions by growth in LB medium at 37 °C to an A₆₀₀ of 0.8, followed by induction with IPTG (0.2 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). Cells were lysed by sonication, and cell debris were removed by centrifugation and filtration (0.45 μ m). The cell lysates were loaded onto a HisTrap 5 mL column (GE), equilibrated in buffer 1. The column was washed extensively with buffer 1 containing 75 mM imidazole, and His-tagged proteins were eluted at 350 mM imidazole. The SUMO-His₆ tagged constructs (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, and ACP_{5a}-ACP_{5b}-DD) were then incubated with His-tagged human rhinovirus 3C protease (1 μ M) 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 containing 20 mM imidazole. Size exclusion chromatography on a Superdex

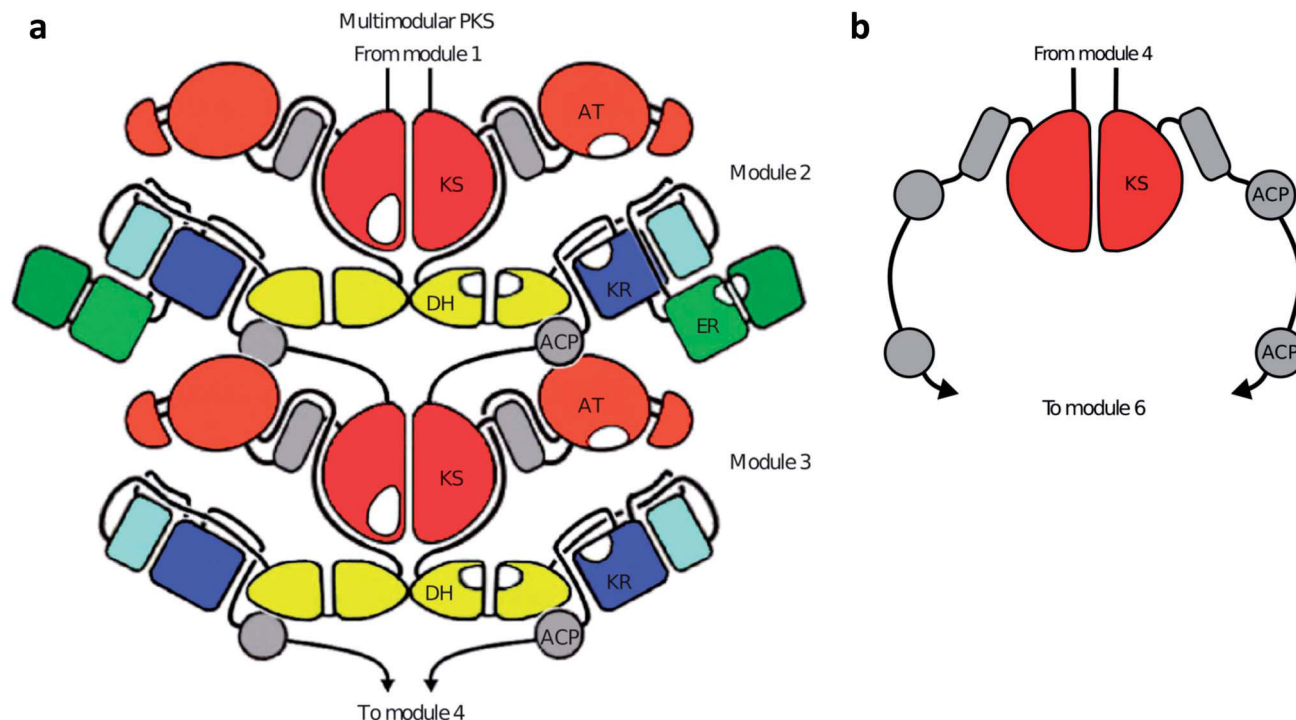


Fig. 7 Structures of PKS modules. (a) Structural model of *cis*-AT PKS modules incorporating various complements of reductive domains, based on the X-ray structure of animal FAS. The overall structure is homodimeric, with the ketosynthase (KS; red) and dehydratase (DH; yellow) domains comprising the dimer interface. The remaining monomeric catalytic domains (acyl transferase (AT; orange), ketoreductase (KR; blue) and enoyl reductase (ER; green)) loop out from the two-fold axis. Thus, each module comprises two independent reaction centers, each of which is accessible by a single acyl carrier protein (ACP) domain (the involved active sites are indicated in white). Each ACP additionally interacts with a downstream KS domain to which, in this example, it is covalently tethered. The grey box between the KS and AT domains represents the intervening, folded linker, a homologue of which is found in virginiamycin module 5 (see Discussion in the text). Artwork reprinted by permission from Macmillan Publishers Ltd: *Nat. Chem. Biol.*, 2012, 8, 615, copyright (2012), and from AAAS (*Science*, 2008, 321, 1315). (b) Schematic of the structure of an *apo trans*-AT PKS module as revealed by SAXS coupled with NMR and homology modeling. The core of the structure is also formed by the KS domain, but together, the two ACP domains, the linker regions and the terminal dimerization/docking domains form an open architecture.

75 26/60 column (GE) in buffer 1 resulted in a homogeneous preparation of each protein (Fig. S2†). Following the HisTrap column, the His₁₀-tagged KS construct was purified to homogeneity by size exclusion chromatography on a Superdex 200 26/60 column (GE) in buffer 1 (Fig. S2†). The His₁₀-tagged constructs KS₅-ACP_{5a}-ACP_{5b} and KS₅-ACP_{5a}-ACP_{5b}-DD were purified by ion exchange chromatography on a HiLoad 16/10 Q Sepharose column (GE) equilibrated in buffer 2 (20 mM sodium phosphate (pH 8)) which was washed with buffer 2 containing 250 mM NaCl. The constructs were then eluted in 400 mM NaCl. Size exclusion chromatography on a Superdex 200 26/60 column (GE) in buffer 1 resulted in a homogeneous preparation of each protein (Fig. S2†). Production of ¹³C,¹⁵N-enriched ACPs (ACP_{5a}, ACP_{5b}, ACP_{5a}-ACP_{5b}) for NMR structure elucidation was carried out by growth in M9 minimal medium. The minimal medium was supplemented with ¹⁵NH₄Cl (0.5 g L⁻¹) and glucose or ¹³C₆-glucose (2.0 g L⁻¹) as the sole sources of nitrogen and carbon.

Correction of the sequence of *virFG*

In order to correct the end of *VirF*, we used a reverse primer complementary to the original end of the gene (5'-

tttccgaagctcTCAGCACCGCGCGCCGG-3'; the capitalized bases correspond to nucleotides 29 985–30 002 of GenBank AB283030.2) and a forward primer complementary to the original beginning of *virF* (5'-aaaaaacctggatgCCGCCG-GAGCGCTGGC-3'; the capitalized bases correspond to nucleotides 32 211–32 226 of GenBank AB283030.2) and sequenced. This analysis revealed a missing C nucleotide at the end of *virF* (original position 30064 of *virF*). Addition of this nucleotide results in a frameshift in the gene which removes the stop codon. *virF* and *virG* are in fact a single gene, which we have designated as *virFG*. To correct the beginning of *virFG*, we designed a forward primer complementary to the upstream *virE* sequence (5'-aaaaaacctggcCCGCCGCTCCGACACCCCTG-3'; the capitalized bases correspond to nucleotides 32 615–32 636 of GenBank AB283030.2), and a reverse primer complementary to the original *virF* gene (5'-tttccgaagcttTCA TGTGACGGGGCGGGCGTTC-3'; the bases in bold correspond to the complement of nucleotides 30 514–30 534 of GenBank AB283030.2). The PCR products were inserted into the pCR-BLUNT vector (Invitrogen) and sequenced. This analysis revealed that a nucleotide (T) had been incorrectly added at position 32 254 of the published *virF* sequence, and that in fact,

the gene begins at nucleotide 32 539. The corrected sequence of *virFG* corresponds much more closely to the comparable region of the homologous pristinamycin cluster (gene *snaE3*).⁴¹ The final sequence of the translated VirFG is provided in Fig. S8.†

Native and denaturing electrospray ionization mass spectrometry

Prior to mass analysis, the samples were subjected to buffer exchange into 50 mM ammonium acetate (pH 6.9), using an Ultrafree-0.5 centrifugal filter device with a cut-off of 10 kDa (Millipore). Mass spectrometry measurements were performed in positive ion mode on a micrOTOF-Q (BRUKER Daltonics) equipped with an ESI source. The injection rate for the samples was 6 $\mu\text{L min}^{-1}$. Spectra were smoothed using the Gaussian method and were baseline subtracted. For native analysis, the desalted sample was diluted into 50 mM ammonium acetate to a final concentration of 20 μM . The ESI needle voltage was set to 4.5 kV, the nebulization gas (Ar) pressure was 1.7 bar, the drying gas flow was 6 L min^{-1} , the source temperature was 200 °C and the capillary exit voltage was set to 198 V. The acquisition range was m/z 2000–10 000. For denaturing analysis, the desalted sample was diluted in $\text{H}_2\text{O-ACN-FA}$ (50 : 50 : 1 volume) at a final concentration of 10 μM . The ESI needle voltage was set to 4.5 kV, the nebulization gas (Ar) pressure was 0.8 bar, the drying gas flow was 4 L min^{-1} , the source temperature was 190 °C and the capillary exit voltage was set to 155 V. The acquisition range was m/z 500–2500.

Small-angle X-ray scattering (SAXS) data collection

The SAXS data were collected on the SWING beamline at the Synchrotron SOLEIL, using an AVIEX170170 CCD detector. Frames were recorded using an energy of 12 keV. The sample-to-detector distance was set to 1899 mm, leading to scattering vectors q ranging from 0.0005 to 0.5 \AA^{-1} . The scattering vector is defined as $q = 4\pi/\lambda \sin \theta$, where 2θ is the scattering angle. The protein samples were injected with an automatic sample changer into a pre-equilibrated HPLC-coupled size-exclusion chromatography column (Bio-SEC 300, Agilent). The online purification system that delivers the eluted fractions into the measurement cell was developed on the SWING beamline.⁶⁰ After equilibrating the column in the protein buffer (50 mM sodium phosphate (pH 7.5), 150 mM NaCl, 5% glycerol), 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 be important for the catalytic activity of *cis*-AT PKS,^{28–30} presumably through effects on the structural integrity of the proteins. Similarly, the buffers used for recent enzymatic analysis of KS domains from *trans*-AT PKS contained both salt and glycerol⁶¹ while that for activity experiments with a *trans*-AT contained phosphate.¹³ 80 μL of the protein sample (at 15 mg mL^{-1}) were then injected and complete data sets acquired. All experiments were performed at 10 °C. In contrast to classical SAXS experiments which are conducted in batch using several protein concentrations within a standard range (e.g. 0.1–10 mg mL^{-1}), the fact that data

collection is coupled to a gel filtration column means that analysis of the required multiple concentrations of the protein occurs within a single experiment (as measured downstream *via* the absorbance at 280 nm with an *in situ* spectrophotometer), as many different positions within the elution peak are sampled during the course of the measurement (typically 50–100 frames are acquired).⁶⁰ Each acquisition frame of the experiment yields a scattering spectrum which is then analyzed by the dedicated, in-house program Foxtrot, yielding a R_g (radius of gyration) as well as an $I(0)$ value (which depends on the protein concentration at that location in the elution peak) from the Guinier law (approximation $I(q) = I(0)\exp(-q^2R_g^2/3)$ for $qR_g < 1.3$). It is thus the fact that the concentration naturally varies across a gel filtration peak that provides a full range of dilutions as used in classical SAXS experiments. Under our experimental conditions (buffer, dilution over the gel filtration step, *etc.*), the fact that we observed a constant R_g for a significant proportion of the concentrations present in our gel filtration peaks means that these measurements were independent of concentration, and thus, that we were effectively under conditions of infinite dilution. In the following step of data analysis, all the frames that exhibited the same R_g as a function of $I(0)$ were averaged, thus ensuring that the data reflected only the signal arising from the protein structure, and not from intermolecular interactions. Processing of the averaged data was performed using the program package PRIMUS.⁶² The distance distribution function $P(r)$ and the maximum particle diameter D_{max} were calculated by Fourier inversion of the scattering intensity $I(q)$ using GNOM.³¹

Ab initio shape determination

The overall *ab initio* protein shapes were calculated from the experimental SAXS data using the bead-modeling program DAMMIN.³¹ At least ten to fifteen independent fits were carried out without imposition of symmetry for the monomeric *apo* ACP_{5a}-ACP_{5b} construct, and with a 2-fold symmetry restraint for the KS₅-linker, *apo* KS₅-ACP_{5a}-ACP_{5b} and *apo* module 5 proteins; in the absence of such an imposed symmetry, the fit between the experimental and model SAXS curves was poorer (greater χ^2). The results of several DAMMIN models were then compared using the program SUPCOMB³³ in order to determine a consensus model. Low-resolution models were then averaged using the DAMAVER and DAMFILT packages.⁶³ In addition, *ab initio* shape determination by a multiphase dummy atom mode using MONSA³⁵ which permitted the simultaneous fitting of scattering data sets obtained on KS₅-ACP_{5a}-ACP_{5b} and KS₅-ACP_{5a}-ACP_{5b}-DD constructs and the individual KS₅-linker domain, allowed for the localization of the domains within the various multienzymes.

In the case of the *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD constructs, the calculation of the *ab initio* molecular forms was carried out without imposition of symmetry ($P1$) because this yielded a better fit to the data than when second order symmetry was imposed ($P2$) ($\chi^2 = 1.21$ vs. 1.6 for ACP_{5b}-DD; 1.3 vs. 1.55 for ACP_{5a}-ACP_{5b}-DD). We attempted to determine using MONSA the individual contributions of the *apo* ACPs to the calculated

molecular forms (that is, to try to place the ACPs into the forms), by fitting the diffusion curves of the two ACPs (calculated from the NMR structures using CRY SOL) and ACP_{5a}-ACP_{5b} in the case of ACP_{5a}-ACP_{5b}-DD, into the diffusion curves obtained on the overall proteins. In our hands, this calculation proved impossible due to the high mutual similarity of the ACPs, their inherent positional flexibility, and to the fact that the inclusion of ACP_{5a} and the ACP_{5a}-ACP_{5b} linker changed the relative orientations and distances between ACP_{5b} and the DD.

Molecular modeling

A three-dimensional structural model of the KS₅-linker was generated by pairwise comparison of profile hidden Markov models using the HHpred server (<http://toolkit.tuebingen.mpg.de/hhpred>).⁶⁴ Local sequence alignment was performed using PSI-BLAST and a multiple sequence alignment including information about predicted secondary structures was produced. The closest homologues to the VirA KS₅-linker within the Protein Data Bank are the DEBS KS₅-AT₅ didomain (40% sequence identity; PDB: 2HG4²⁴) and a didomain KS-branching (B) enzyme from the subunit RhiE (37% identity; PDB: 4KC5¹⁵). Three-dimensional models of the KS₅-linker were then constructed using either PDB 2HG4 or 4KC5 as template, using MODELLER.³⁴ The resulting models were then dimerized by superposition onto the KS domains of the dimeric crystal structures from the PDBs 2HG4 and 4KC5. Manual inspection of the resulting dimeric models revealed no steric clashes within the models. CRY SOL⁴³ was then used to calculate the theoretical scattering curve of the two homology models. The quality of the 3D modeling was determined using the discrepancy χ^2 , defined according to Konarev and colleagues.⁶²

The theoretical scattering curves predicted by CRY SOL⁴³ from the DEBS-derived and RhiE-derived models of the VirA KS₅-linker gave an equally good fit to the experimental data ($\chi^2 = 1.581$ vs. 1.582, respectively) (Fig. 3c). Part of the reason that the RhiE KS did not produce a superior model may be that its KS-AT linker region contains a 23-residue insertion relative to VirA, which is also not present in DEBS module 5 (Fig. S3†). Correspondingly, analysis of the RhiE KS-B structure shows that the KS-AT linker incorporates an additional β -strand relative to the DEBS KS-AT linker region.

Having selected the model based on the DEBS KS₅ template, the atomic model was superimposed onto the *ab initio* calculated shape of the protein using SUPCOMB.³³ Rigid body modeling against SAXS data was performed using SASREF.³⁵ In particular, SASREF was employed to position the KS and the linker with respect to each other by simulated annealing, while minimizing the discrepancy between the SAXS experimental and calculated curves. Figures displaying the protein structures were generated using PYMOL (Version 1.3, Schrödinger, LLC) and the low-resolution bead models from DAMMIN were converted into mesh.

Analysis of the relative positions of ACP_{5a} and ACP_{5b} was also attempted using SASREF. However, ten independent rounds of calculation produced divergent results that yielded poor fits to the experimental data for angles between 0.0005 and 0.2 Å⁻¹,

with χ^2 between 25.0 and 25.3. Furthermore, the R_g values calculated for the models generated by SASREF are smaller than those derived from Guinier and the $P(r)$ function. The failure of SASREF to more accurately place the two domains relative to one other reveals that the SAXS data cannot be accounted for by a single conformation of the tandem of ACP domains.

NMR data acquisition

To minimize the amount of protein needed, 300 μ L of protein solution was loaded into a 4 mm NMR tube and the tube was inserted into a 5 mm NMR tube containing 150 μ L D₂O.⁶⁵ The *apo* ACP_{5a} and ACP_{5b} domains were analyzed at a concentration of 1.0 mM and the *apo* ACP_{5a}-ACP_{5b} didomain at a concentration of 0.3 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 ¹⁵N-¹H and ¹³C-¹H HSQC spectra and 3D HNCACB, CBCA(CO)NH and HNCO experiments (reviewed in ref. 66). Aliphatic side chain resonance assignments were obtained from 3D (H)CC(CO)NH, H(CCCO)NH and HCCH-TOCSY experiments. ¹H chemical shift assignments of aromatic side chains were based on 2D heteronuclear (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE and 2D homonuclear COSY, TOCSY and NOESY experiments. Tautomeric states of histidines were determined using 2D heteronuclear long-range ¹⁵N-¹H HSQC.⁶⁷ To collect NOE-based distance restraints for the ACP_{5a} and ACP_{5b} structure calculations, 2D ¹H homonuclear NOESY experiments were carried out on unlabelled samples of *apo* ACP_{5a} and ACP_{5b}, while 3D ¹⁵N NOESY-HSQC and ¹³C NOESY-HSQC were recorded on uniformly ¹³C, ¹⁵N-enriched *apo* ACP_{5a} and ACP_{5b} 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 (<http://www.cgl.ucsf.edu/home/sparky/>)).

Chemical shift changes

To express the chemical shift variations of the corresponding amide groups in *apo* ACP_{5a} vs. *apo* ACP_{5a}-ACP_{5b} and *apo* ACP_{5b} vs. *apo* ACP_{5a}-ACP_{5b}, the chemical shift change was defined as $\Delta\delta_{av} = (\{\Delta\delta H^2/2\} + \{\Delta\delta N^2/10\})^{1/2}$.⁶⁸ The most significant changes were observed for Arg6734, Gln6735, Gly6811 and Ala6812 of ACP_{5a} (Fig. 5). Consideration of the *apo* ACP_{5a} structure strongly suggests that the changes can be attributed to the presence of the linker in the didomain. In the ACP_{5a} structure, the β -methyl group of Ala6812 is located in proximity to the Arg6734-Gln6735 amide (measured distance less than 4 Å). However, when the linker is present (residues 6811-6828), Ala6812 is displaced from its position near Arg6734 and Gln6735, affecting not only its chemical shift, but those of the two previously neighboring residues. The chemical shift of Gly6811 is altered by the change in covalent linkage. This interpretation, in combination with the fact that the chemical shifts of ACP_{5b} are essentially

unchanged in the didomain, leads us to conclude that the two *apo* ACPs behave as independent domains.

NMR structure calculation

The first round of structure calculation was performed with CYANA 3.⁶⁹ 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.⁷⁰ In each cycle, 100 structures starting from random torsion angle values were calculated with 10 000 steps of torsion angle dynamics-driven simulated annealing. 1968 NOE-based distance and 114 backbone angle restraints were used for structure calculation of *apo* ACP_{5a}, while 1846 NOE-based distance and 114 backbone angle restraints were used for *apo* ACP_{5b}. The angle restraints were obtained from ¹³C_α, ¹³C_β, ¹³C', ¹⁵N, ¹HN and ¹H_α chemical shifts using TALOS+,⁷¹ with an assigned minimum range of ±20°. 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 rMD simulations in Amber 11⁷² using the generalized Born solvent model. Possible atom clashes within CYANA structures were regularized by a 1 ps energy minimization in the Amber force field, and then 20 ps of rMD (restrained molecular dynamic) 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 mol⁻¹ Å⁻² and 50 kcal mol⁻¹ rad⁻², respectively. The final *apo* ACP_{5a} and ACP_{5b} representative ensembles correspond to the 20 conformers from each calculation with the lowest restraint and Amber energy terms. The quality of the structures was assessed using PROCHECK-NMR⁷³ and the MolProbity⁷⁴ server. The Ramachandran statistics for *apo* ACP_{5a} and ACP_{5b} are respectively: most favored regions, 99.1% and 91.9%; additionally allowed regions, 0.9% and 7.6%; generously allowed regions, 0.0% and 0.5%; and disallowed regions, 0.0% and 0.0%. MolProbity scores (percentiles) for the lowest restraint energy structures of ACP_{5a} and ACP_{5b} are 0.99 (100th) and 0.88 (100th), respectively.

Sequence analysis

The sequences of the analyzed *cis*- and *trans*-AT PKS⁴ were accessed *via* PubMed, or in the case of the class 2 C- and N-terminal docking domains, from ref. 46. 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). The boundaries of the KS domain, and the KS-AT and post-AT linkers were assigned by comparison with the sequences of homologous regions from the erythromycin polyketide synthase,²⁴ while the boundaries of the structured ACP domains were determined by

NMR (this study). The putative dimerization/docking regions were assigned as the sequences lying downstream from the last ACP domains of the respective subunits.

Accession numbers

Chemical shifts for *apo* ACPs 5a and 5b of VirA module 5 have been deposited in the Biological Magnetic Resonance Bank with accession codes 19 536 (ACP_{5a}) and 19 539 (ACP_{5b}). Protein Data Bank coordinates have been deposited under accession codes 2MF4 and 4CA3, respectively.

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 University of Lorraine (UL), and the Lorraine Region (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 and to the Bruker MicroTOF-Q (Service Commun de Spectrométrie de Masse et de Chromatographie, UL) was greatly appreciated. We thank Dr J. Pérez and Dr P. Roblin for help during data collection on the beamline SWING at the synchrotron SOLEIL (GIF-sur-YVETTE), and Prof. P. F. Leadlay, Dr B. Wilkinson and Dr V. Receveur-Bréchet for helpful comments on the manuscript.

Notes and references

- 1 S. Smith and S. C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.
- 2 J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, **29**, 1111–1137.
- 3 L. Katz, *Methods Enzymol.*, 2009, **459**, 113–142.
- 4 J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.
- 5 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.
- 6 S. Sudek, N. B. Lopanik, L. E. Waggoner, M. Hildebrand, C. Anderson, H. Liu, A. Patel, D. H. Sherman and M. G. Haygood, *J. Nat. Prod.*, 2007, **70**, 67–74.
- 7 C. M. Thomas, J. Hothersall, C. L. Willis and T. J. Simpson, *Nat. Rev. Microbiol.*, 2010, **8**, 281–289.
- 8 N. Pulsawat, S. Kitani and T. Nihira, *Gene*, 2007, **393**, 31–42.
- 9 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936.
- 10 A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
- 11 J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, 2012, **8**, 615–621.
- 12 T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, **321**, 1315–1322.
- 13 F. T. Wong, X. Jin, I. I. Mathews, D. E. Cane and C. Khosla, *Biochemistry*, 2011, **50**, 6539–6548.
- 14 A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hothersall, E. Płoskoń, P. Wattana-amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi,

- J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas and M. P. Crump, *Nat. Chem. Biol.*, 2013, **9**, 685–692.
- 15 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher and C. Hertweck, *Nature*, 2013, **502**, 124–128.
- 16 M. Czjzek, H.-P. Fierobe and V. Receveur-Bréchet, *Methods Enzymol.*, 2012, **510**, 183–210.
- 17 R. Dehn, Y. Katsumaya, A. Weber, K. Gerth, R. Jansen, H. Steinmetz, G. Höfle, R. Müller and A. Kirschning, *Angew. Chem., Int. Ed.*, 2011, **50**, 3882–3887.
- 18 N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth and C. Hertweck, *Chem. Biol.*, 2012, **19**, 1164–1174.
- 19 W. Mattheus, L.-J. Gao, P. Herdewijn, B. Landuyt, J. Verhaegen, J. Masschelein, G. Volckaert and R. Lavigne, *Chem. Biol.*, 2010, **17**, 149–159.
- 20 O. Erol, T. F. Schäberle, A. Schmitz, S. Rachid, C. Gurgui, M. El Omari, F. Lohr, S. Kehraus, J. Piel, R. Müller and G. M. König, *ChemBioChem*, 2010, **11**, 1253–1265.
- 21 M. A. Matilla, H. Stöckmann, F. J. Leeper and G. P. C. Salmond, *J. Biol. Chem.*, 2012, **287**, 39125–39138.
- 22 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *ChemBioChem*, 2010, **11**, 1840–1849.
- 23 C. T. Calderone, *Nat. Prod. Rep.*, 2008, **25**, 845–853.
- 24 Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11124–11129.
- 25 Y. Tang, A. Y. Chen, C.-Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 931–943.
- 26 Y.-Q. Cheng, J. M. Coughlin, S.-K. Lim and B. Shen, *Methods Enzymol.*, 2009, **459**, 165–186.
- 27 A. M. Haapalainen, G. Meriläinen and R. K. Wierenga, *Trends Biochem. Sci.*, 2006, **31**, 64–71.
- 28 M. Bycroft, K. J. Weissman, J. Staunton and P. F. Leadlay, *Eur. J. Biochem.*, 2000, **267**, 520–526.
- 29 R. Pieper, S. Ebert-Khosla, D. Cane and C. Khosla, *Biochemistry*, 1996, **35**, 2054–2060.
- 30 K. E. H. Wiesmann, PhD Thesis, University of Cambridge, 1995.
- 31 D. Svergun, *J. Appl. Crystallogr.*, 1992, **25**, 495–503.
- 32 H. Fischer, M. de Oliveira Neto, H. B. Napolitano, I. Polikarpov and A. F. Craievich, *J. Appl. Crystallogr.*, 2010, **43**, 101–109.
- 33 M. Kozin and D. Svergun, *J. Appl. Crystallogr.*, 2001, **34**, 33–41.
- 34 N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper and A. Sali, *Curr. Protoc. Protein Sci.*, 2007, **50**, 2.9.1–2.9.31.
- 35 M. V. Petoukhov and D. I. Svergun, *Biophys. J.*, 2005, **89**, 1237–1250.
- 36 R. N. Perham, *Annu. Rev. Biochem.*, 2000, **69**, 961–1004.
- 37 U. Trujillo, E. Vázquez-Rosa, D. Oyola-Robles, L. J. Stagg, D. A. Vassallo, I. E. Vega, S. T. Arold and A. Baerga-Ortiz, *PLoS One*, 2013, **8**, e57859.
- 38 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker and V. N. Uversky, *J. Mol. Biol.*, 2006, **362**, 1043–1059.
- 39 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, *Biochim. Biophys. Acta*, 2010, **1804**, 996–1010.
- 40 P. E. Wright and H. J. Dyson, *Curr. Opin. Struct. Biol.*, 2009, **19**, 31–38.
- 41 Y. Mast, T. Weber, M. Gözl, R. Ort-Winklbauer, A. Gondran, W. Wohlleben and E. Schinko, *Microb. Biotechnol.*, 2011, **4**, 192–206.
- 42 L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angew. Chem., Int. Ed.*, 2011, **50**, 2795–2798.
- 43 D. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.
- 44 C. Bich, S. Baer, M. Jecklin and R. Zenobi, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 286–289.
- 45 M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Comput. Biol.*, 2007, **3**, 1827–1835.
- 46 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, **20**, 1340–1351.
- 47 E. J. Brignole, S. Smith and F. J. Asturias, *Nat. Struct. Mol. Biol.*, 2009, **16**, 190–197.
- 48 T. J. Buchholz, T. W. Geders, F. E. Bartley, 3rd, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.
- 49 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117–125.
- 50 I. M. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
- 51 S. E. A. Ozbabacan, H. B. Engin, A. Gursoy and O. Keskin, *Protein Eng., Des. Sel.*, 2011, **24**, 635–648.
- 52 P. Caffrey, D. J. Bevvitt, J. Staunton and P. F. Leadlay, *FEBS Lett.*, 1992, **304**, 225–228.
- 53 P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner and R. Kolter, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 305–310.
- 54 Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili and X. S. Xie, *Science*, 2010, **329**, 533–538.
- 55 A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399–6408.
- 56 I. Thomas, C. J. Martin, C. J. Wilkinson, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2002, **9**, 781–787.
- 57 K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzer, B. K. Rubio, S. J. Robinson, P. Crews and J. Piel, *Nat. Chem. Biol.*, 2009, **5**, 494–501.
- 58 G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist and C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 139–144.
- 59 J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2013, **8**, 1263–1270.
- 60 G. David and J. Pérez, *J. Appl. Crystallogr.*, 2009, **42**, 892–900.
- 61 M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chem., Int. Ed.*, 2013, **52**, 1143–1147.
- 62 P. V. Konarev, V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun, *J. Appl. Crystallogr.*, 2003, **36**, 1277–1282.
- 63 V. Volkov and D. Svergun, *J. Appl. Crystallogr.*, 2003, **36**, 860–864.

- 64 J. Söding, A. Biegert and A. N. Lupas, *Nucleic Acids Res.*, 2005, **33**, W244–W248.
- 65 M. W. Voehler, G. Collier, J. K. Young, M. P. Stone and M. W. Germann, *J. Magn. Reson.*, 2006, **183**, 102–109.
- 66 J. Cavanagh, W. Fairbrother, A. Palmer and N. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press Inc., New York, 1996.
- 67 J. G. Pelton, D. A. Torchia, N. D. Meadow and S. Roseman, *Protein Sci.*, 2008, **2**, 543–558.
- 68 L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli and K. J. Weissman, *Chem. Biol.*, 2010, **17**, 705–716.
- 69 P. Güntert, *Methods Mol. Biol.*, 2004, **278**, 353–378.
- 70 T. Herrmann, P. Güntert and K. Wüthrich, *J. Mol. Biol.*, 2002, **319**, 209–227.
- 71 Y. Shen, F. Delaglio, G. Cornilescu and A. Bax, *J. Biomol. NMR*, 2009, **44**, 213–223.
- 72 D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
- 73 R. A. Laskowski, J. A. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton, *J. Biomol. NMR*, 1996, **8**, 477–486.
- 74 V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 12–21.

Supplementary Information

Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module

Jack Davison¹, Jonathan Dorival¹, Hery Rabeharindranto^{1,2}, Hortense Mazon¹, Benjamin Chagot^{1,*}, Arnaud Gruez^{1,*} & Kira J Weissman^{1,*}

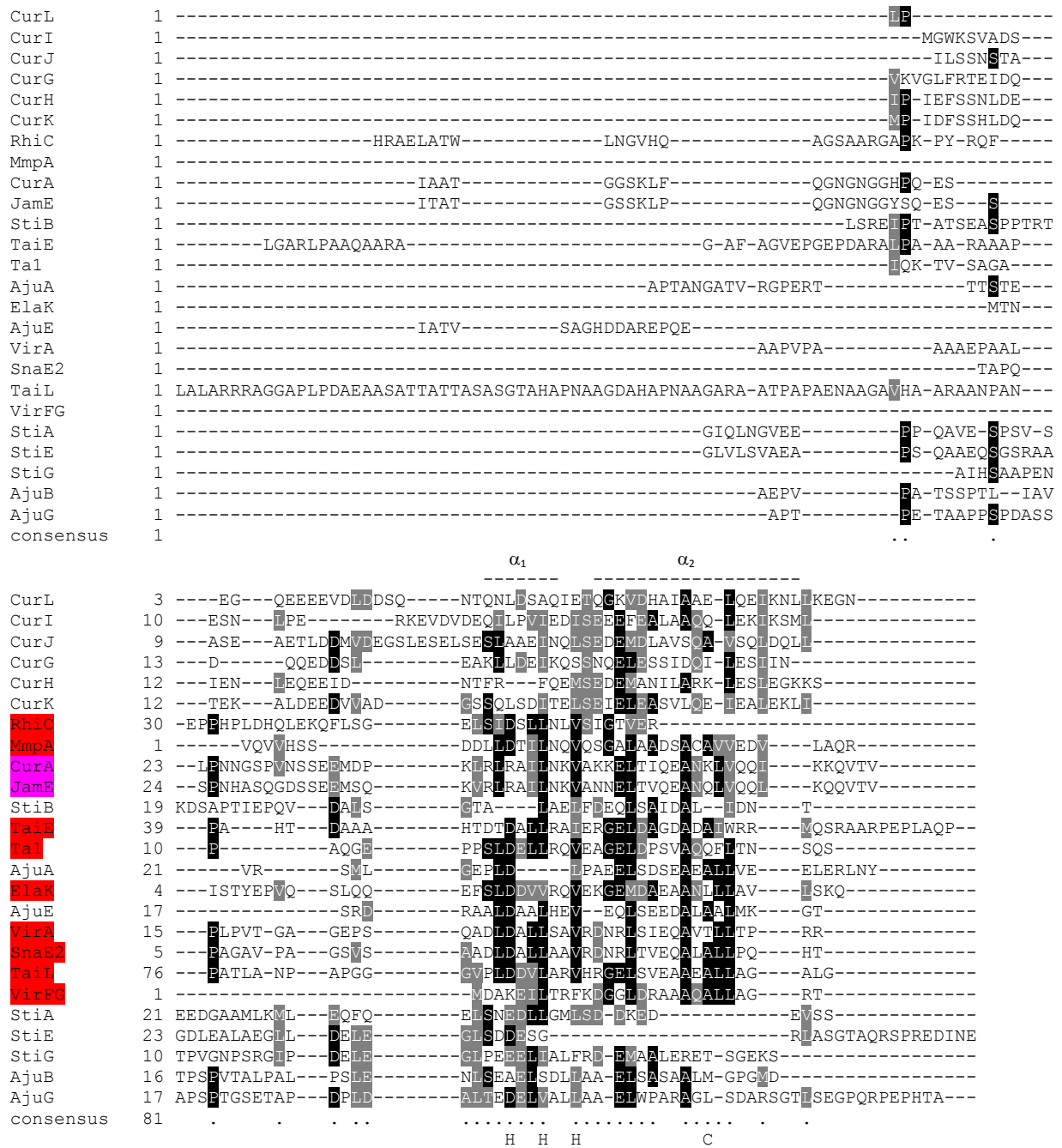
¹Molecular and Structural Enzymology Group, Université de Lorraine, IMoPA, UMR 7365, Vandœuvre-lès-Nancy, F-54500, France. CNRS, IMoPA, UMR 7365, Vandœuvre-lès-Nancy, F-54500, France.

²Present address: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés – INSA, UMR INSA/CNRS 5504 – UMR INSA/INRA 792, 135 Avenue de Rangueil, 31077 Toulouse CEDEX 4, France.

*Correspondence should be addressed to B.C. (benjamin.chagot@univ-lorraine.fr), A.G. (arnaud.gruez@univ-lorraine.fr) or K.J.W. (kira.weissman@univ-lorraine.fr)

Supplementary Figures

a



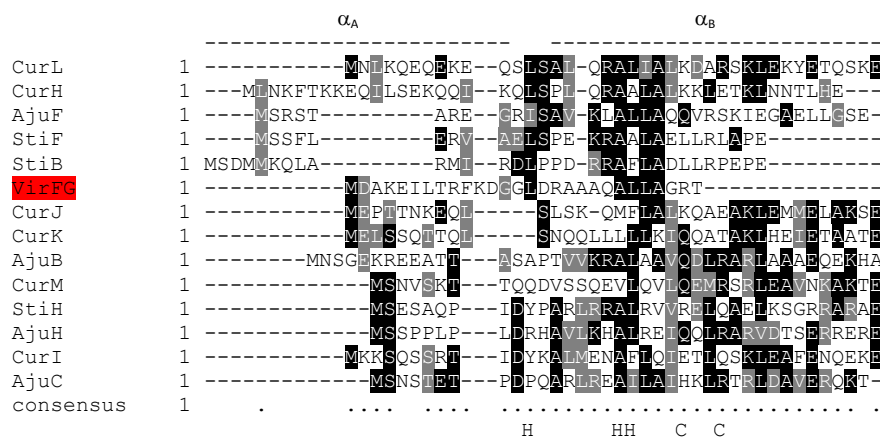
b

Fig. S1 Sequence analysis of putative dimerization/docking regions in various PKS. (a) Multiple alignment of the C-terminal regions of PKS subunits from *trans*-AT PKS (in red), with those of *cis*-AT PKS¹ (black, except for those of subunit CurA and the closely-related JamE (in pink)). The sequences from the *trans*-AT PKS (with the exception of VirFG) shown lie downstream of the conserved sequences of single (Ta1) or tandem acyl carrier protein (ACP) domains associated with β -modification reactions, which are located at the ends of subunits. The VirFG sequence represents the N-terminus of module 6, which we propose to be a docking element interacting with the C-terminus of VirA. 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 α -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.¹ (b) Multiple alignment of the N-terminal regions from the partner *cis*-AT PKS subunits to those shown in (a), alongside that of VirFG (in red). The positions of the two α -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.¹ Key: Cur, curacin (*cis*-AT PKS); Rhi, rhizoxin (*trans*-AT); Mmp, mupirocin (*trans*-AT); Jam, jamaicamide (*cis*-AT); Sti, stigmatellin (*cis*-AT); Tai, thailandamide (*trans*-AT); Ta, myxovirescin (antibiotic Ta; *trans*-AT); Aju, ajudazol (*cis*-AT); Ela, elansolid (*trans*-AT); Vir, virginiamycin (*trans*-AT); Sna, pristinemycin (*trans*-AT).

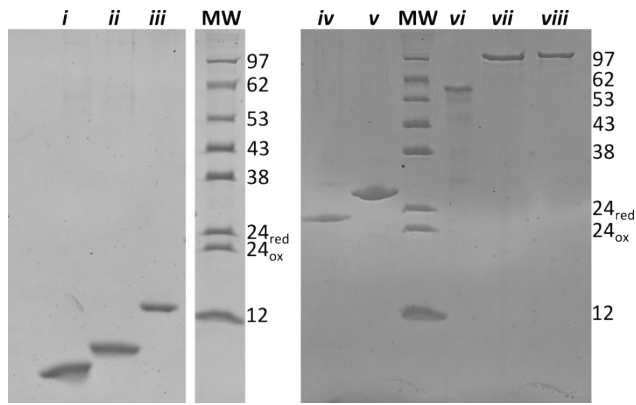


Fig. S2 SDS-PAGE analysis of protein preparations used in this study. Key to lanes: i) ACP_{5a} (calc'd: 8.9 kDa); ii) ACP_{5b} (calc'd: 9.2 kDa); iii) ACP_{5b}-DD (calc'd: 14.3 kDa); iv) ACP_{5a}-ACP_{5b} (calc'd: 19.4 kDa); v) ACP_{5a}-ACP_{5b}-DD (calc'd: 24.5 kDa); vi) KS₅-linker (calc'd: 62.2 kDa); vii) KS₅-ACP_{5a}-ACP_{5b} (calc'd: 98.5 kDa); viii) KS-ACP_{5a}-ACP_{5b}-DD (calc'd: 103.6 kDa). The molecular weights of the markers are indicated.

KS

VirA_Module5 1 **LIATIGVAGRFPF_AD_L_DAIW_NL_GRD_ITEVP_DRWD_HAI_YD_DK_IVDRTH_KWGGEVD_VE_FD_AIE_ISPREAC_IVDPO_RLFLIS_MAL**
SnaE2_Module5 1 **AVATIGVAGRFPFQAGL_DELRNLIAGRSITENRRRWDAEALDPDRSRVDRYTWGGGFCQVNFPAPIFS_ISPREACIVDPO_RLFLIACAAAL**
RhiE_Module11 1 **YIATIGLGGYFPAISL_DELMONLANVYCSDEPADRWDSKPKYKRVLCRTICIN_SFKIKVYKPYSHFRKPKVYADHMSPEVRLFLQVHTFT**
MmpA_Module7 1 **DAITIGLAGRMPQARTLHEFMQVLSQGRDCEITPDRWRHSRYSDADAPGRTYARWGGFTDGVAFEPAPFFSISPRAMAMPQRLFLQTAHEAT**
BaeL_Module6 1 **DTAITIGSGRMPQANTLCEFWNLSSEDCITETIPDRWRHSLYDADDKDKGRTYKRWGGFTKDVYKDFDPPFFSISPRAKIMDPO_RLFLQCVYETM**
PksL_Module6 1 **EITAITIGSGRMPQANTLDEWKRDRGRDCEITPDRWRHSLYDADDKDKGRTYKRWGGFTKDVYKDFDPPFFSISPRAKIMDPO_RLFLQCVYETM**
EtnE_Module6 1 **ATAITIGSGRMPQANDLAEWENLRSKDCITETIPDRWRHRYADABGKPKYSKRWGGFTDGLDFFDPPFFSISPREACIMDPO_RLFLLCAYQTL**
EtnD_Module1 1 **DVAIVGVGRFPQARGP_HOLWENLRSGRCAITAVPDRWSD-----AAGEFRGGFVDADCFRAFFKISPAEIMDPO_RLFLLEVYHTL**
DifL_Module14 1 **PIAITIGSGRFAESNTLHEWELHLAGRDLTKKADR---WD---V-----SDQCQVSTEDDPCFPVFFHSIGREATYMDPO_RLFLDESCKAL**
LnmJ_Module5 1 **ELAITIGLAGRFPSSADTDEWEHLAGVDRDRPDRKDTAIRAN-----PAIRELRGGTDSVLTFRARLEISPRAMAMPDPO_RLFLQTAWRVF**
DEBS_Module5 1 **PIAITIGMAGRFPDVSDFSEWFFVSGGALAEERDRGW-----PDPDALGGMIAAGDPLACFFSISPREALAMPDPO_RLFLMIEISWEAL**

VirA_Module5 100 **DDAGYTPDRLVSAEDPVHRRDVGVEVGMVYCYQHHEAEERLRGNPV_LANSAVMSIANRVSYFFDFGSPSVAVDTACSSALTAVHLAVESLRGTSE**
SnaE2_Module5 100 **DDAGYTPDRLVSTADPVHRKDKVVEVGMVYCYQHHEAEERLRGNPV_LANSAVMSIANRVSYFFDFHOGPSVAVDTACSSALTAVHLAVESLRGTSE**
RhiE_Module11 100 **EDAGYKRTILSR-----YNDVGYVLLGYSNHYHYGFENFVSNV-RGSM-AS_SGMALIMTVSYVGLTGPSFEDDMCSSSHTLHACQSLRHFDETK**
MmpA_Module7 100 **EDAGYTRGCLAASARHEDAEANVGVVGVYFQYQYHGAQQTAEERF-VLMSFS_SSIANRVSYVNGCHSPSMALDAMCSSITLHACQSLRHFDETK**
BaeL_Module6 100 **EDAGYTRKTEKSGDLLANVGVVGVYFQYQYHGAQQTAEERKSL-ALITGNPSSIANRVSYVNGCHSPSMALDAMCSSITLHACQSLRHFDETK**
PksL_Module6 100 **EDAGYTRHGRKRDALDLSGVVGMVYFQYQYHGAQQTAEERSL-ALITGNPSSIANRVSYVNGCHSPSMALDAMCSSITLHACQSLRHFDETK**
EtnE_Module6 100 **EDAGYTRHGRKRDALDLSGVVGMVYFQYQYHGAQQTAEERSL-ALITGNPSSIANRVSYVNGCHSPSMALDAMCSSITLHACQSLRHFDETK**
EtnD_Module1 90 **EDARVKKPSSLE-----ERKLVGVNVCNDYAEHLHEDAYDLSYGSTSTSCSLSNRVSYFFDFRGPSTVDTACSSSLAVHLARALRAGECE**
DifL_Module14 87 **EDAGYACLSIR-----REGCVLAGSCGGYCAIFKQGGPA-QA---FWNHNSTVTPARAVHLNLOGPAITVDTACSSSLTAHACQSLRHFDETK**
LnmJ_Module5 92 **EDAGYRPADLA-----CAPGCLFVGVATHYDLDLKNVGA-VQAHTATIAHVS_SIANRVSYVNGCHSPSVAVDTACSSSLTAHACQSLRHFDETK**
DEBS_Module5 90 **ERACHPDVSIR-----SATGVETGVGVYGRPDEAPDE-VLGVYGTASSVASGRVAVCLGLEGPAITVDTACSSSLTAHACQSLRHFDETK**

VirA_Module5 198 **VAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
SnaE2_Module5 197 **VAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
RhiE_Module11 193 **MYLAGGLNIMYHYHTVNTSQNFTSITSEVNSGCVADGTVIGEGCAVLLKPLRALDADDCIYGVKIGAMNAGEBNGNVPNEDLQTLAROAM**
MmpA_Module7 198 **VAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
BaeL_Module6 197 **AAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
PksL_Module6 197 **AAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
EtnE_Module6 194 **LAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
EtnD_Module1 182 **AALAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
DifL_Module14 175 **MVAAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
LnmJ_Module5 183 **LAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**
DEBS_Module5 181 **LAIAGGVNV_IHP_KYFNLQG_F_SSLG_RCRRSFGAGGGDGYVPGEGVAVLLKPLRAL_DGDHI_AVIRGTA_NHGGRTNGYVFNPAQA_LV_KAL**

VirA_Module5 297 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
SnaE2_Module5 296 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
RhiE_Module11 293 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
MmpA_Module7 297 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
BaeL_Module6 296 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
PksL_Module6 296 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
EtnE_Module6 293 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
EtnD_Module1 281 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
DifL_Module14 274 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
LnmJ_Module5 282 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**
DEBS_Module5 280 **RDAGLHEDLYIEAHGTCT_LGDPTEI_GLA_AF_ABDGTEIETIKG_ESSILTGSVKSNVGH_BSAAG_AL_KVLLQLRHR_L_PSLHANE_NF_T**

KS-AT linker RhiE

VirA_Module5 389 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
SnaE2_Module5 388 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
RhiE_Module11 384 **DFADIPFVVOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
MmpA_Module7 388 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
BaeL_Module6 387 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
PksL_Module6 387 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
EtnE_Module6 386 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
EtnD_Module1 379 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
DifL_Module14 365 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
LnmJ_Module5 380 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**
DEBS_Module5 371 **DFGVPF_VOQ_LA_NRWHDG-----AAE_F_RAGISSFGAGG_NAHLIVEE_PAPAPRP_ATGRRG_QVFLSARTTAAAYARDL_DHL_RARSAGE**

KS-AT linker DEBS 3

VirA_Module5 482 **BPSAADVAITFAVGRVDRARRAVPADLILLAGELAEGRPRVPPP-----AEEI_AAMISER**
SnaE2_Module5 480 **BPRADITAFHFAVGRPAIERRAALAADSLALASHEBALAAGRPRVWR-----T-----EAAADDPVRLWIDEG**
RhiE_Module11 478 **BAPAQIAYTLOVGRNRRNLA_RCR_RQA_SRA_N_CID_HYOSSAD-----SKI----F--YRFQES_DAVQP_ESDL-NDPLAPLTONLND**
MmpA_Module7 484 **EPDLQAIAYTLOVGRQAWEMRALLVITGDELREDFRIDCALEGSPSWSGCLPEAHSL---A--T--RPSEQACALRQMFQDADLGRRRWQGE**
BaeL_Module6 485 **EADLSRIAYTLOVGRQAWEMRALLVITGDELREDFRIDCALEGSPSWSGCLPEAHSL---A--T--RPSEQACALRQMFQDADLGRRRWQGE**
PksL_Module6 485 **EADLSRIAYTLOVGRQAWEMRALLVITGDELREDFRIDCALEGSPSWSGCLPEAHSL---A--T--RPSEQACALRQMFQDADLGRRRWQGE**
EtnE_Module6 483 **QRDLPSIAYTLOVGRQAWEMRALLVITGDELREDFRIDCALEGSPSWSGCLPEAHSL---A--T--RPSEQACALRQMFQDADLGRRRWQGE**
EtnD_Module1 473 **PVALSELAYLICTREBFEERLALREDSAAI_SQDRFVCS_VDPAGCFR-GNVRTAKGAPVDA-----VDAHIGGRALSSAELMRYCO**
DifL_Module14 450 **DSLDGNFTYLLGRKHHRFRACVSADELRKRVSEWLVKRELPGVFV-SNLKQKPAEEAGMKTFGMECTEQCRSAASPAQYRENLEHADLCREY**
LnmJ_Module5 470 **EADLSRIAYTLOVGRQAWEMRALLVITGDELREDFRIDCALEGSPSWSGCLPEAHSL---A--T--RPSEQACALRQMFQDADLGRRRWQGE**
DEBS_Module5 457 **RNSLAD_GFTLDRASAMR_R_AVVVGRDE_LACLAVADR_RIADR-----GOGPNSPR**

post-AT linker DEBS 3

VirA_Module5 543 **EIDGSPAADAR-----VELVYVPSQRVWLP**
SnaE2_Module5 545 **RDRDREAACGAASVGRRLPLPHYPERVRCWYDLOIAHLHRQGLSAGQEPAFARDHLDFGRAPSDREQAMPGTAPAPAAVPSAAEGAQVS--APTGV**
RhiE_Module11 562 **SOVDASAY-----DPRISLEAYEKRWYEEGESSL**
MmpA_Module7 576 **ANDSPHHS_HREARGLPTYPBARQRYMAPVESVSVASVA-----QGGK-----RRLD-----TLPV**
BaeL_Module6 579 **PENNDKIQM-QRQKRRISLPAVPAARDRYIDVNAKAEKRETEFF-----AF--VO-----IAAAPE-----PV**
PksL_Module6 579 **PENNDKIQM-QRQKRRISLPAVPAARDRYIDVNAKAEKRETEFF-----AF--VO-----IAAAPE-----PV**
EtnE_Module6 577 **TDWDDQPPAGVRRRLSPPYPPAMKRYVGPWAPSAAPPA-----T-----AF--SR--SAPPTS-----IAGADAA--VRSDAASP**
EtnD_Module1 559 **SRFRGRLAG-AVERTALPLSSEBRRYVTRQVAPAEVSVT-----VAPPGAAIEKPS-----GARRAPGEERFAAPAVSSG**
DifL_Module14 549 **DLPETRLDG-QQYVTPPLPAPYPLERYWAADVNSLPGQAMKSE-----TQT-VLSSPCEHEGL-----SL-----SPLTEIE**
LnmJ_Module5 547 **DLPETRLDG-QQYVTPPLPAPYPLERYWAADVNSLPGQAMKSE-----TQT-VLSSPCEHEGL-----SL-----SPLTEIE**
DEBS_Module5 547 **DLPETRLDG-QQYVTPPLPAPYPLERYWAADVNSLPGQAMKSE-----TQT-VLSSPCEHEGL-----SL-----SPLTEIE**

Long linker connecting KS and B domains, RhiE


```

VirA_Module5 640 -PARAAVRPA-VPAPTAVPTL-TLRG---AAEPARRPATDPMEKPM--TSDRKISLRPLAPATAATRPAAPA-TPATPAAAPV-----
SnaE2_Module5 644 -PRRVEPAPALVPRPAPASAP-AAPG---AVTPPATDRPDPQETPFM--TRDRKISLRPLGTAPTAPAPAPAPAPVSVAAAPVPAATTPAAAPAPVA
RhiE_Module11 -----
MmpA_Module7 636 -----A-----DS-----A-----PIVPPA-----PR--LSQAPLAAKA-----
BaeL_Module6 637 -----IE-----VT-----T-----LHEETVKEKPSNTTLOLTSQAKPKRTV-
PksL_Module6 630 -----IP-----KP-----S---V-DREASGKPNITLQSLMTNQDRLER-
EtnE_Module6 649 -----IA-----RSAPKERAARADAAPISLVALVPPAAMDT---AASSAARSAPVSLRAEPASAIV-
EtnD_Module1 636 VEHTVSL---SPVGAAVNQAAVAGSAMGAAPYGGGSPAGGAASSA---IAGV-----ERVTPRATEGVGGHMPGTTATARRRAS-
DifL_Module14 621 MT-----A-----FE-----LQKFTVLOGLDCTVSV-
LnmJ_Module5 621 AS-----ATAPTREPVPAP-----PAP-----APAPATTAPQVPRADV-
DEBS_Module5 -----

```

ACP

```

VirA_Module5 717 APPAAAPAAVPAAGRQETIIVLA VLYLELD E FL LGVDSILGVFEV V N Y V V AT LYD PT A IA GAPAPAP--
SnaE2_Module5 737 APSVAAPVAGAVDFVAARVAEEI GEVLAGNLYMEPADIDFQSFQMLGMSISMEFTALNNAKALCDKRAEYHPPTAAFRHVAATGGTTPPPAPA
RhiE_Module11 -----
MmpA_Module7 660 -----APVVPVADDECAQFTRQSLAAMLYCPGQGRDGSRFLHLGLDSVIAAQWIREINKHYQKPKPDGCTYTVFKAFPTQWVTCQOPTQATA--
BaeL_Module6 669 ---QF--ASRPVAVSAKVSAVTESLADVLYMDADDIDADTFIDIGMSITGLEWIKSNKAYGTSITVTKYVYPTIRQRFAPLQKEGTRPILEEA
PksL_Module6 661 ---V--PSDTEETITAFALCDEITAGLAENLYMDQNEIDPEAFIDIGMSITGLEWIKSNKAYGTSINVTKYVYPTTRDSEVYVHEHSTQAGEKKQ
EtnE_Module6 709 ---ARRASDEPPASGGEPITGRFRASLARALCTDEADIDQAVFVNLGLDSVIGVEWARSEREELDVPTASTIYHPHTLRRLSAHVSGLTSARADLRA
EtnD_Module1 712 ---VDNAAGGRSLAAEQEITGSLAERALFMRADVEPRFEMELGDSVIGVEWVAVNNTVYCLSPATQVYHPHTLRRLSAHVSGLTSARADLRA
DifL_Module14 645 ---QPEKAVSEDNRTLEERTSLADLLEISPEIDAEPEFIDIGMSITGLEWIKSNKAYGTSINVTKYVYPTTRDSEVYVHEHSTQAGEKKQ
LnmJ_Module5 654 ---PPPALPEAAAATVVGLELGEISKULGSPSEETENAPFGLGLDSIYRMELVRLNEAEGCDKRAEYHPHTIIGKLTFFVPLIIGPAGAAPAA
DEBS_Module5 -----

```

ACP

```

VirA_Module5 815 -----TFAPAVPASPAPV-----AVAVDPAVARALREGLAATLYEPCDIDDEASFNLGLDSITGVFEV
SnaE2_Module5 837 -----PHATTPAPAAPAATAPA-----NA-----APVAAGERTAAULETIRECLAQTLVQVWDITTEATFNNSLGLDSITGVFEV
RhiE_Module11 -----
MmpA_Module7 750 -----APVQREPVATAP-----Q-----PGAQASARESVDYIKQSTFELTFLPGQTRSGAQFTDIDIGMSITGTONM
BaeL_Module6 763 -TEP-----QKTEAVKKPKNIMLQPLNAHQHVQIQAADILEMSETAASAEPEEPLEAHDLESKARADVLYMRRHEVTEIEAFIDIGMSITGHEWH
PksL_Module6 757 -TETYTPTRQKTVVPAKPNANISLQPLEHHQPVQEEAETTQVAAREISASRQYTVAIETHEHRESADVLYMRRHEVTEIEAFIDIGMSITGHEWH
EtnE_Module6 806 APGPAPVQ-----PEAPASPALP-----ARGAQHEISIPVAVRPVEEVVAGLITQDLAAALYALASDIPRLVYVHGLGDSITGVFEV
EtnD_Module1 807 HRRPRG-----GAAEA--GRARADLGAEEGLTSSLAALYALASDIPRLVYVHGLGDSITGVFEV
DifL_Module14 740 ETAVRP-----QIPEA-ETPLAGISDGTTRKELKDSLADILRLKPPDIEHEAFTIEMGLDSITGVFEV
LnmJ_Module5 749 EPVMAE-----APQQS-SASLEDLVQDVEREELG-----RTAPAKSFTVNFSGSFDMLRVV
DEBS_Module5 -----

```

```

VirA_Module5 876 FVNDYGGDDEKAGLYDPSLAPLRRVAER
SnaE2_Module5 907 AFHNNAYGLDERAGVLYDHESSAALAAHITSR
RhiE_Module11 -----
MmpA_Module7 814 RGVSRHHSIQLAADALVYTWPTIKSLADEVDRR
BaeL_Module6 856 KANNRYGCTDCNVTKVYDPTIRQFADLRAQ
PksL_Module6 856 KANNRYGCTSFVTRVYDPTIRQFADLRAQ
EtnE_Module6 884 RRNNATHLRLKRTVLEPHPTIHLAERHATL
EtnD_Module1 868 RANNHAYGLSLPACTVYDHPPTIGRLARHLSFM
DifL_Module14 802 QSNNTYQASITENLWYVYPTIATLAGYLTGS
LnmJ_Module5 800 ASSEFVVGALRRKTLRPHPTIGALAAHIAET
DEBS_Module5 -----

```

Fig. S3 Sequence analysis of a selection of modules (type KS-ACP-ACP and KS-ACP-ACP-ACP) from *trans*-AT PKS showing the homology to the KS-AT and post-AT linker regions of a typical *cis*-AT PKS module. The sequence of the KS-AT and post-AT linkers from module 5 of the erythromycin (DEBS) *cis*-AT PKS, for which structures are known,² are shown within and above the alignment (in blue and green, respectively), while the corresponding regions from the recently-published KS-B didomain structure of RhiE³ are shown in light blue and olive. Comparison of the KS-AT linker regions between the VirA, RhiE and DEBS modules shows that the RhiE sequence contains a large insertion. The positions of the conserved domains (based on the structure of the KS₅-AT₅ didomain of DEBS² and the NMR structures of VirA ACP_{5a} and ACP_{5b} (this work)) within the modules are also indicated in color. Key: Dif, difficidin; Vir, virginiamycin; Sna, pristnamycin; Mmp, mupirocin; Bae, bacillaene from *Bacillus amyloliquefaciens*; Pks, bacillaene from *B. subtilis*; Lnm, leinamycin; Etn, etnangien. In each case, the subunit name and module number are indicated.

a

ACP

```

MmpB_Module8_ACP-ACP-ACP_2 1 AILLDEIVALLCOQLRTVAGIDPHPTDLHDGFDSSLLTQLLAQSSSYGVLEDFEGSVLEDAIACGLVQVQQRHGAEPASLRVPQVQ--ERRA-----
MlnB_Module2 1 SEKSTLREITGVLKTEERFFDPISEYGFDSIMFTMGELNETLGLSMPDDEGLTDFNLELEKSEYSEHLGCSFSSDE--SAA-----
MmpB_Module8_ACP-ACP-ACP_1 1 DPSDILRGLYCOQLKVAQEDDTTAFSDYGFDSVMLTETATLNRITLLELSTAALEPFTIQALAAHQOGARTAESQPP--APG-----
BaeJ_Module2 1 KWHGALHLLSSLLKYGQDELDLDELSEYGFDSVFTVPTGLNEANQKLAFTLFEHGLTRSLAAHLLTDEAEAG--PS--QPEEK---HTA-----
MlnH_Module11 1 TITDELSSLEGEATNTPSNKARHQSVEYGLDSSKQLHHSKMLFEELDTQLVPTLLFNNHHDLAALBEKQDVRNTPAFAAFKFKTQTKEP-----S
AlbI_Module1 1 SPLDQVKKLLGRLLKMDVDTLDSHPLEYGVDSIVAETAMALRETPG-FEVSLEFETQSIDLLLESEEQAP---LLA-----T-
XabB_Module1 1 SPLDQVKKLLGRLLKMDHATLDSHPLEYGVDSIVAETAMALRETPG-FEVSLEFETQSIDLLLESEEQAP---LLA-----T-
DifI_Module1 1 KSVDFLKKVIGCALKVPAPHRLLAHPLEIAYGLDSILVVRITNALKQVFGS-VSSTLFEFYQITIDLADVFLSRRKAEERQFG--NGRAHPENA---S-
DifI_Module6 1 KTAAEFLKAAAKVLLKMPASEENLAPLLELYGLDSILVVRITNALKQVFGS-VSSTLFEFYQITIDLADVFLSRRKAEERQFG--NGRAHPENA---S-
MlnF_Module8 1 ITEGDIQQLIADIKGTAFETISVVKGVYEGGLDSTLQAVKALREKICRELVPTLLFEYPTIQKLANFEEGKETSVEKD----LPLH---YEN
ChiC_Module5 1 AKARLRQLVAGRIGRDEFSVPTIRGEVYEGGLDSTLQAVKALREKICRELVPTLLFEYPTIQKLANFEEGKETSVEKD----LPLH---YEN
ChiC_Module4 1 RALACIRBELSARELLEAHEHNDNTRFEPCVDSILLGLAKREKIVYCKLFESEIFFEYPTIRARLAQLSSVKKPAPAPSA--AALAARFPASACHA
LkcC_Module2 1 AVEGFREBTEAGTRRLVYSEVECDPFRMGDSLSSELKFLREYCY-EPATVFEYPTIRLVEELVTEHAAMAESFG--AATATATA-----
DisA_Module2 1 AVEERLIDSLASTLQDRSRKSSDVEPFTTEGDSIFAVEAGVYGRLESLDTRTALENYPTARAAHEHTAATAPSEARPAR--APEFAAQPREQL-PS
MmpA_Module7 1 ECAQRLRQSLAAWLYCFGCHRDGSRFTEGLDSTVIAQWIRREINNKYQOKPADGTYTFYVFKAFPTQVVGTCQVPTQATAAP--VOREPVAT-----
OnnB_Module2_ACP-ACP-ACP_1 1 QVQRQLRRSLAEEELHMTFDMVEDETEYVKMGDSIIAVSWVKINQAGCALGATVYVYTNLDDLOHFPEIAPKPSSTIP--EPELAVSS---
OnnB_Module 1 DWHVLRQLLAEELHMTFDMVEDETEYVKMGDSIIAVSWVKINQAGSEATVYVYTNLDDLOHFPEIAPKPSSTIP--EPELAVSS---
VirA_Module5 1 I E V R L L A V L L P P D L D E F G V D S I L G V A V N Y V V T L Y D P T A A T A S G A P -----
OnnI_Module7 1 QNTQLLSQLAQTLEVPENCITLQVDFSYGVDSILGVNFTQINDEGLGEMNTVIFPHHTHHHTHTTRVYYSRRL-----CLS
RhiC_Module9 1 AITAVALNGLAKTLQLEPFRIMTDFEAYGVDSILGVSEFGLLEALGIELNTALFLFYPTLDSVLDLGGQPAESARLAG--GQGDVDESANGLSAA
consensus 1
  
```

ACP

```

MmpB_Module8_ACP-ACP-ACP_2 93 -----PQR--VVT-----DAEPEQRADRQSPVVVLLSAGSAAQDROVADNLLHIDGQADFDVLDHTLARVTQAGSGMLPVRL
MlnB_Module2 92 -----KDR--REN-----HIE-----GAKMSSLLKDRKEDEPTTF-----SRY
MmpB_Module8_ACP-ACP-ACP_1 85 -----LT--RAQ-----VAQ-----GVRVWAEALKRLEDIGDDDF-----SDY
BaeJ_Module2 89 -----DKS--LQT-----LHT-----ATAMVSGLLKDRREDITLDEL-----SEY
MlnH_Module11 95 EKPVLV---ESLYLQKEW-----TESPLKSEVPEKENTLIFDFDSSSD-----EIR-----KYI
AlbI_Module1 78 -----PAPQDDLQQLKQLARLTKKIDTQDTSKTL-----ESY
XabB_Module1 78 -----PAPQDDLQQLKQLARLTKKIDTQDTSKTL-----ESY
DifI_Module1 93 -----RVQAPAE---SAPIPNVRERGITVTEHLGKTKLKKIPSGRDPVSVL-----TAY
DifI_Module6 93 -----SLSLEKSTEVYKQVSAALLKMPEDDQVHTL-----EYI
MlnF_Module8 91 HTAL-----TYLTHWRQKA-----AAWDPNC---GNKKRCTLIFGYDRDFHMRN-----REN
ChiC_Module5 97 -----GA---AGALD-----TPDTLVLDTPGTFATAPEDALRRKLAGLGGPPAQRSDSGC-----YEL
ChiC_Module4 99 -----AA---AAR-----EGSATLQWVEGQLAGLVKMPAAEDASTRI-----BEL
LkcC_Module2 91 -----A---VA-----SEAGAVVRGAVEGFRVPLAGVHLIDSEVDEDFR-----BEM
DisA_Module2 98 SPQPAPGAPPR---ER--QAT-----SQVQAP-APERPP-PPQVGAQQVRLQALGALAEVMAIDREPSDAIL-----AEC
MmpA_Module7 92 -----PAP-----DFGCAASAQRESQDYDLKSLLELLEDPGQSRGACF-----DEL
OnnB_Module2_ACP-ACP-ACP_1 92 SDSDIYPGFKP-IALQETVAAKKTGNANGSSLAHEKQKPTPAGRPEAIGQDLKDVHVKRQLLAEELHMTFDMVEDETEYVKMGDSIIAVSWVKINQAGSEATVYVYTNLDDLOHFPEIAPKPSSTIP--EPELAVSS---
OnnB_Module 85 -----QSA-----NAVGEPE-----EKAVAMABRQLLAEELHMTFDMVEDETEYVKMGDSIIAVSWVKINQAGSEATVYVYTNLDDLOHFPEIAPKPSSTIP--EPELAVSS---
VirA_Module5 78 -----APAPTFAVAVPASAVFAV-----AVDPAVVARLRELARLYCEPGDIDEAFC-----NTL
OnnI_Module7 82 NAPQLQEVV-----DHGELHRLVSHLAKLLDWAESTTEGDPVF-----SDY
RhiC_Module9 99 HTPPEVSAHTPELVLAELDLPVVT-----AQVMQASQEAALAVGLASALQDLAAQIDNEQEF-----SDY
consensus 101
  
```

```

MmpB_Module8_ACP-ACP-ACP_2 164 GLSVSSLEAEQ-----FRAYLOTIE-HSAYAQVGNALVWASVNR
MlnB_Module2 127 GDSITFTTRGGR-----INQOYCFELPSVFEFCNLAEMTDYVIEQ
MmpB_Module8_ACP-ACP-ACP_1 119 GDSVSSVQITGL-----LNEAFDCLADDTFOAGNWWELTAAADI
BaeJ_Module2 124 GDSVSTVFTVQ-----LNEAYQLELAFIIFEHGHSGLAGYLAIE
MlnH_Module11 141 GSENVYAEAAA--YEMVNEKHYKFNFI---QSAHYSRLLEDKKS
AlbI_Module1 115 GDSIVLIEEANA-----FRYYP-SLDAQMELSLIPRLVAQWQAT
XabB_Module1 115 GDSIVLIEEANA-----FRYYP-SLDAQMELSLIPRLVAQWQAT
DifI_Module1 140 GLDSILVVRITNA-----FRWVED-QMSTLFEFYQITIEEDYVILS
DifI_Module6 129 GDSILVOLTNR-----FRDVP-DISSLFEFYPTGGLAEYVSS
MlnF_Module8 140 GFENTQITLIVSGNRFSYHGQGVYCFHR----EAHFRLAALIKR
ChiC_Module5 152 GLSDLLSARE-----EALLGEOYFVLFVLFHPTLAVHLRRE
ChiC_Module4 138 GIDPTVTSRHH-----EASEEVALDPAOLFHHFLRRLCGMIDEA
LkcC_Module2 136 GDSISMEIKP-----FRVYV-VPATVFEYPTVRLVEFTVE
DisA_Module2 166 GHDACQAVVYVS-----NQALTSATAMDLRCCPLADVFHLLAS
MmpA_Module7 135 GDSVGTQWRG-----VSRRESHCLADATVTPPLKSLADQWDR
OnnB_Module2_ACP-ACP-ACP_1 176 GLDSIIVSWIQ-----LNQAYGLSEATVYVYTNLDDLOHFPEIAPKPSSTIP--EPELAVSS---
OnnB_Module 130 GLDYVMAGSWVQ-----LNQAYGLSEATVYVYTNLDDLOHFPEIAPKPSSTIP--EPELAVSS---
VirA_Module5 132 GLDSILVEIVAF-----VNOYGLDCACTLYDHPSLAALSRHVAG
OnnI_Module7 124 GLDSILGVNFTQ-----ENDDLSGEMNTVIFHTSNALADHNSKT
RhiC_Module9 162 GLDSILGASVDE-----LNEALGHEISALFLFYPTVTLSTVLDH
consensus 201 *
  
```

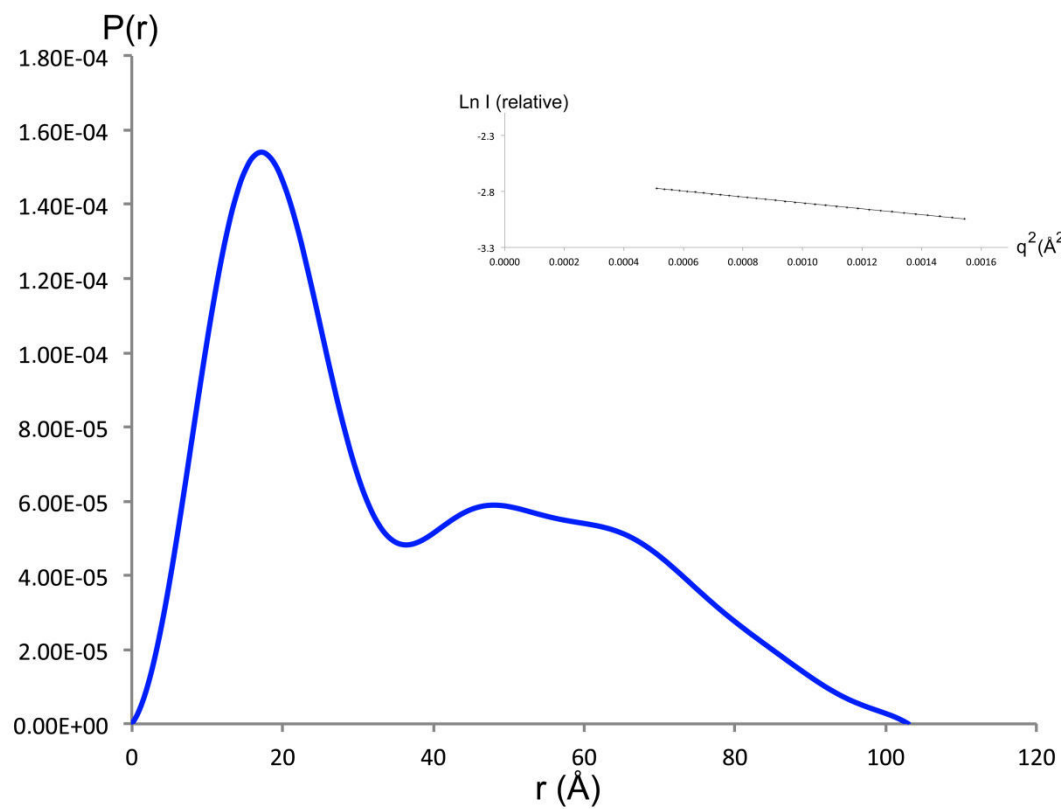
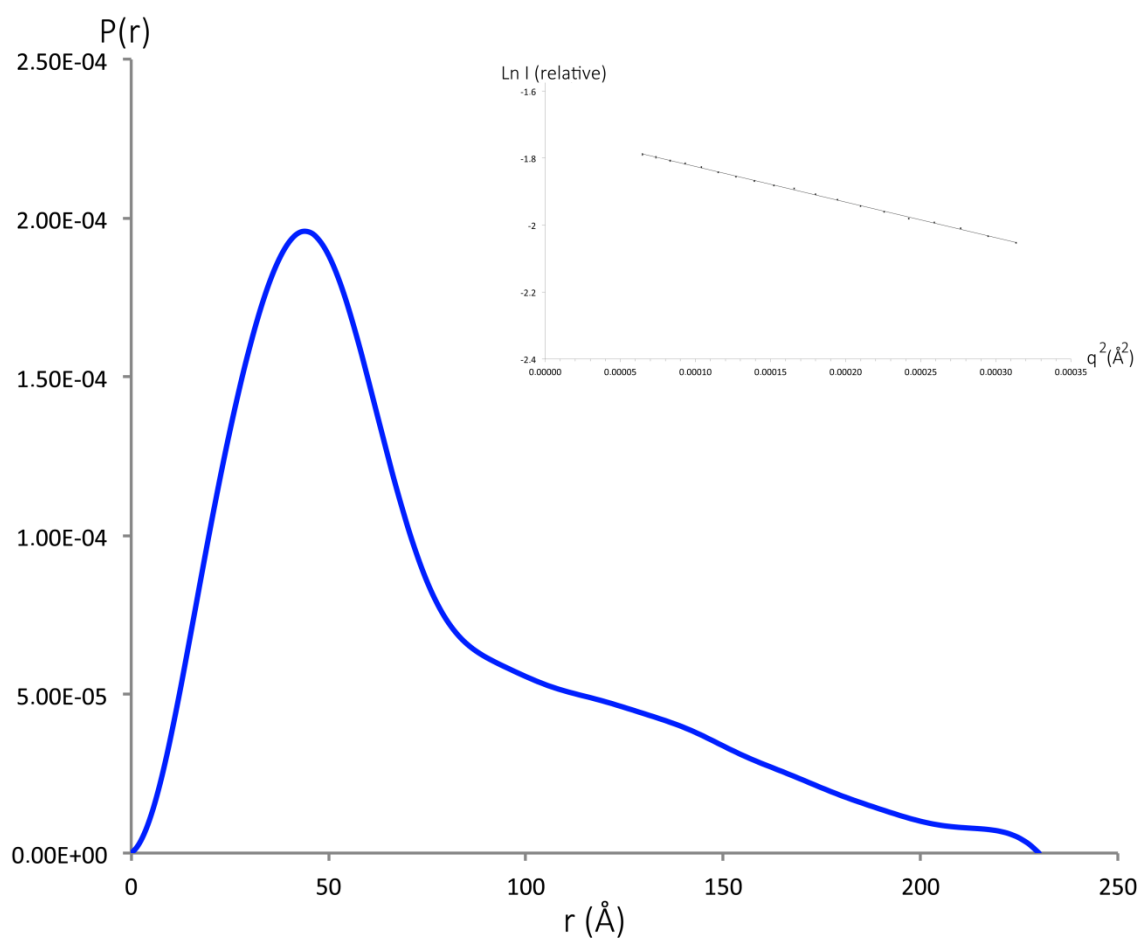
b

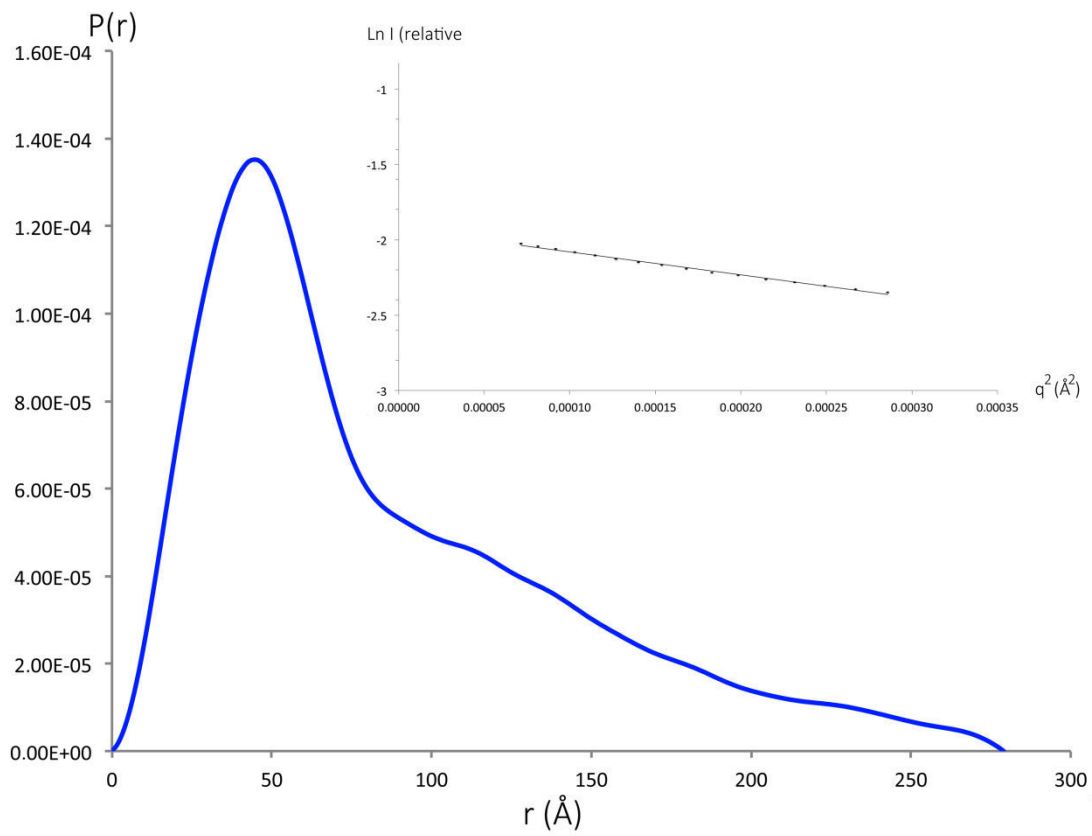
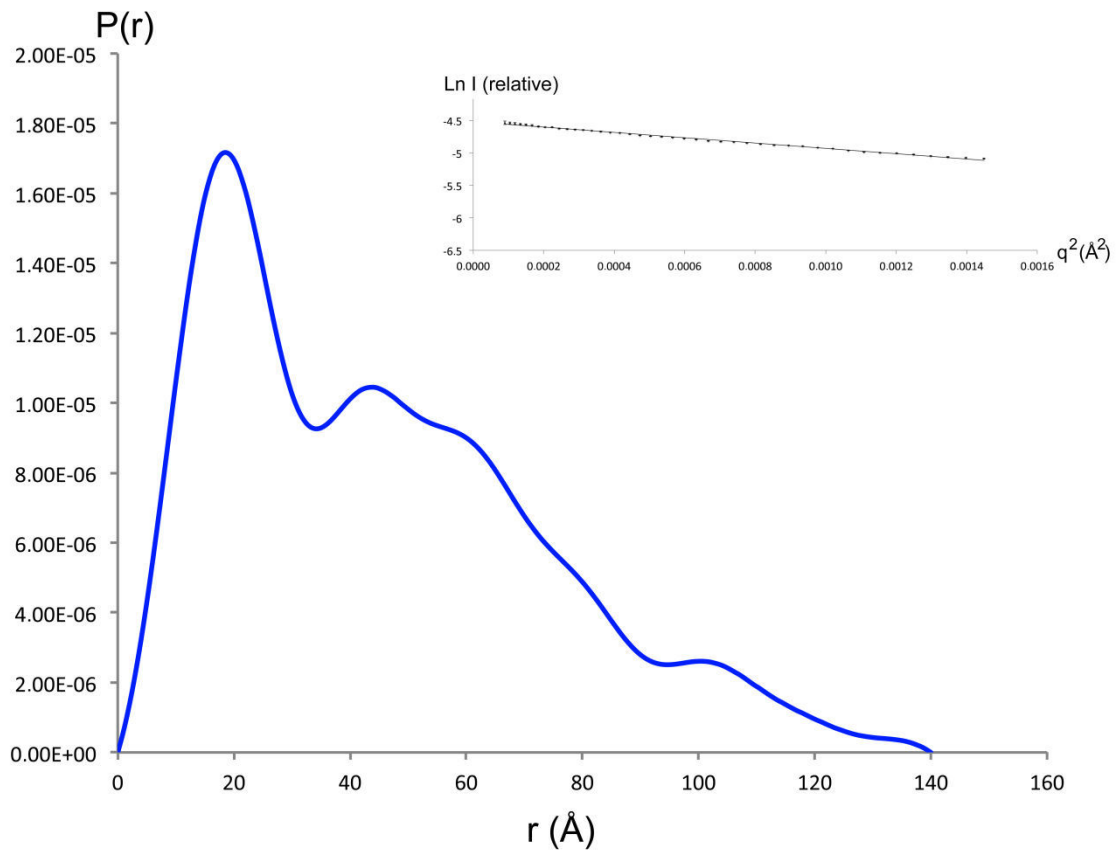
```

M LEVVAEKTGY PTEMLDLDMD MEADLGIDSI
1261 KRVEILGTVQ DQMPNLPELS PEDLAECRTL REIVTYMNSK MPAAAAASAPV TSASNGLDAA
1321 QVQSTMLEV ADKTGYPTM LDLEMDMEAD LGIDSIKRV ILGTVDQQLP TLELSPEDL
1381 AECRTLGEIV SYMNSKLPAA SAVAAPVASA PVASANGLD AAQVQTM TMLD VVADKTGYPT
1441 EMLDLAMDME ADLGIDSIKR VEILGTVDQD LPGLPELNPE DLAEACRTLGE IVDYMNSKLE
1501 AVSTQNVAIQ TAAPVASASN GLDAAQVQGT MLEVVDKGTG YPAEMLD FAM DMEADLGIDS
1561 IKRVEILGTV QDQLPGLPEL NPEDLAECRT LGEIVDYMNS KLPAASTQNV AVQTVAAAPVA
1621 TAPATNGLDA AHVQNTMMNV VADKTGYPAE MLDLAMDMEA DLGIDSIKRV EILGTVDQQL
1681 PGLPELNPE DLAEACRTLGEI VAY
  
```

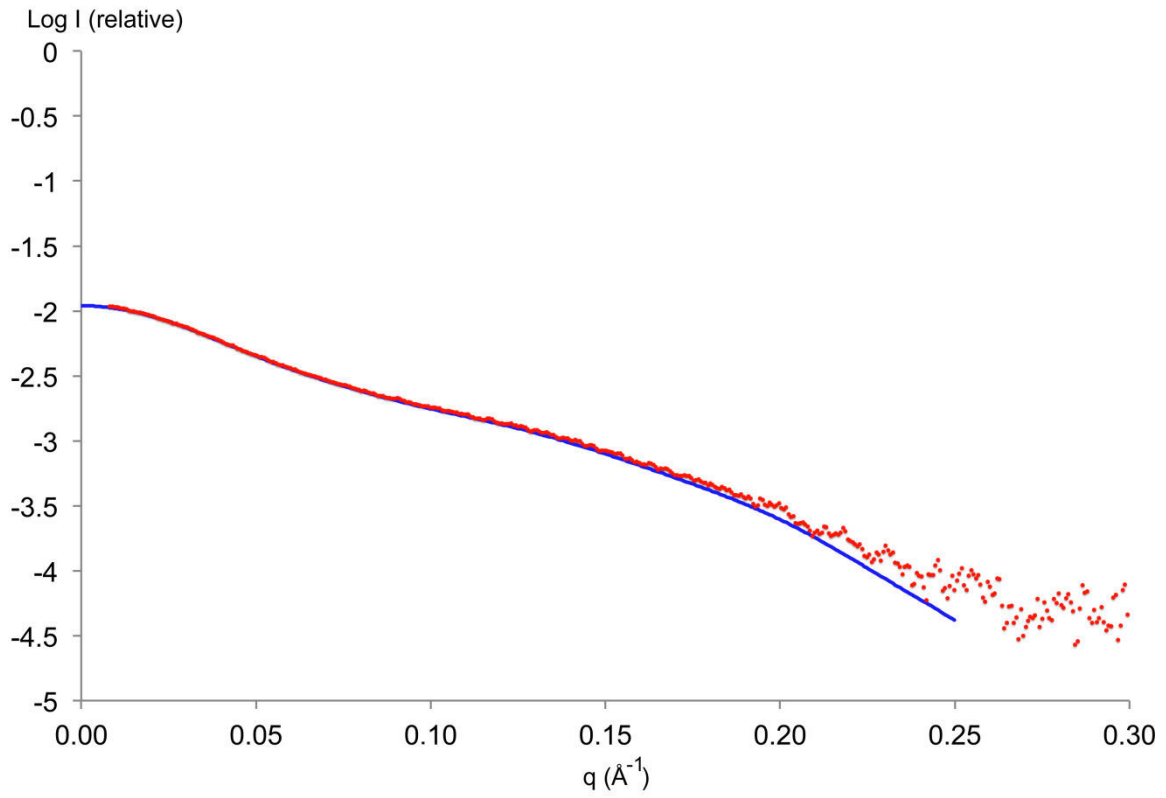
- Linker 1 = 24 residues (P+A = 10; 42%)
- Linker 2 = 29 residues (P+A = 14; 48%)
- Linker 3 = 31 residues (P+A = 10; 32%)
- Linker 4 = 34 residues (P+A = 13; 38%)

Fig. S4 Sequence analysis of ACP-ACP regions in *trans*-AT PKS and PfaA. (a) Sequence alignment of the ACP-ACP regions from a selection of *trans*-AT PKS modules. In each case, the subunit name and module number are indicated. The positions of the conserved ACP domains (based on the presence of secondary structure elements in the NMR structures of Vir ACP_{5a} and ACP_{5b}) are shown for the sequence of VirA. The ACP-ACP linkers of VirA and mupirocin are highlighted in green and red, respectively. Key: Mmp, mupirocin; Mln, macrolactin; Bae, bacillaene from *Bacillus amyloquefaciens*; Alb, albicidin from *Xanthomonas albilineans* Xa23R1; Xab, albicidin from *Xanthomonas albilineans* Xa13; Dif, difficidin; Chi, chivosazol; Dis, disorazol; Lkc, lankacidin; Onn, onnamide; Vir, virginiamycin; Rhi, rhizoxin. (b) Sequence of the tandem ACP region of the PUFA synthase, PfaA (GenBank: AAL01060.1). The positions of the four ACP-ACP linkers are indicated in red, while their respective lengths and content in Pro and Ala are given.

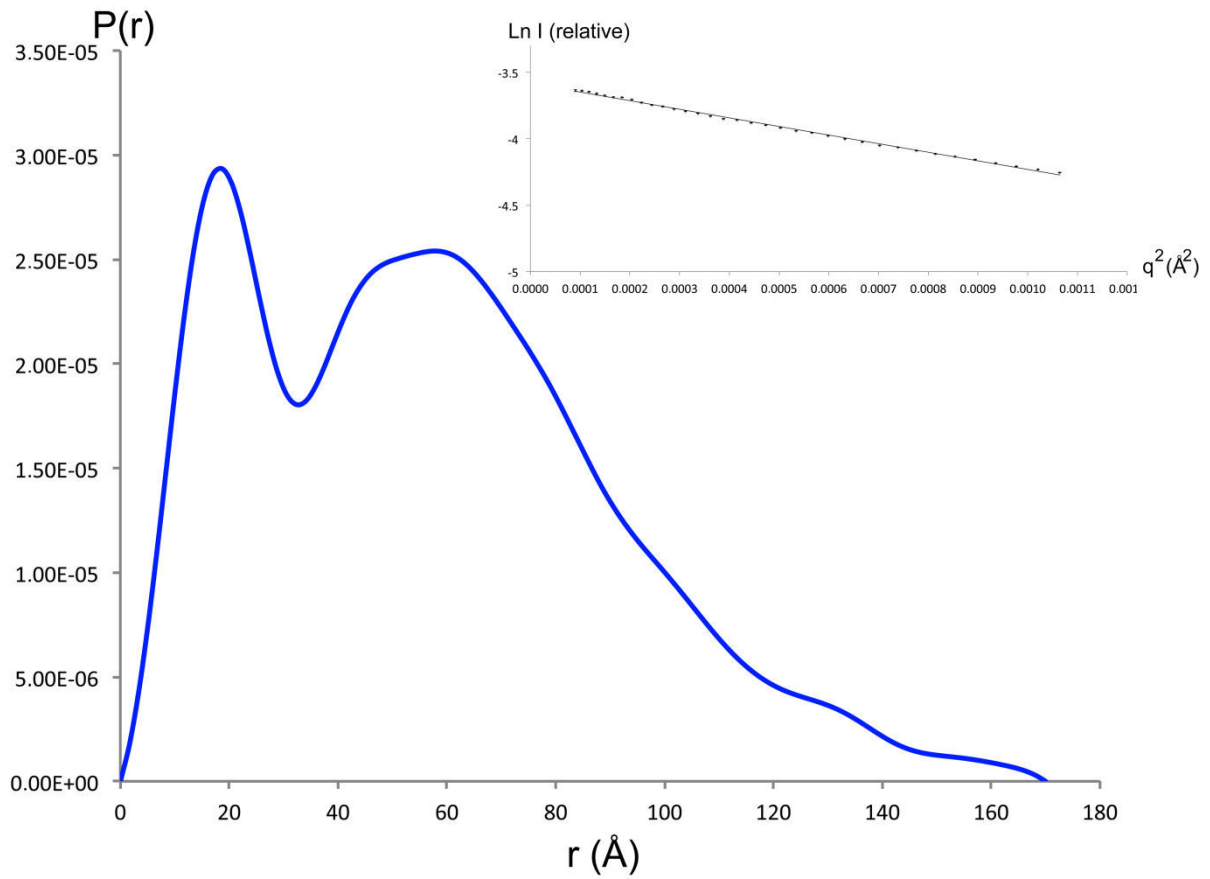
A**B**

C**D**

E



F



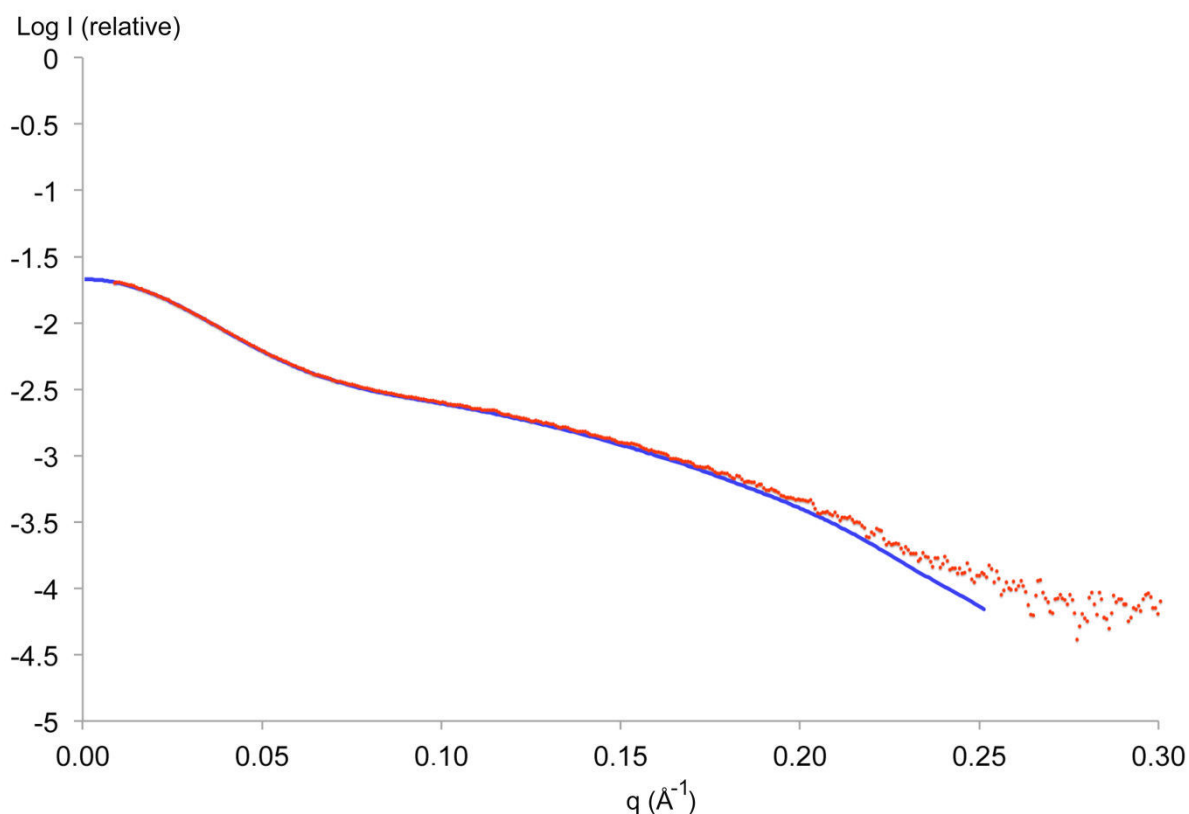
G

Fig. S5 SAXS analysis of the various constructs. (a) The distance distribution function derived for the ACP_{5a}-ACP_{5b} construct calculated with GNOM. Inset is the Guinier plot. (b) The distance distribution function derived for KS₅-ACP_{5a}-ACP_{5b} calculated with GNOM. Inset is the Guinier plot. (c) The distance distribution function derived for KS₅-ACP_{5a}-ACP_{5b}-DD calculated with GNOM. Inset is the Guinier plot. (d) The distance distribution function derived for the ACP_{5b}-DD calculated with GNOM. Inset is the Guinier plot. (e) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data acquired on the ACP_{5b}-DD (red dots). (f) The distance distribution function derived for the ACP_{5a}-ACP_{5b}-DD calculated with GNOM. Inset is the Guinier plot. (g) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data acquired on the ACP_{5a}-ACP_{5b}-DD (red dots).

a

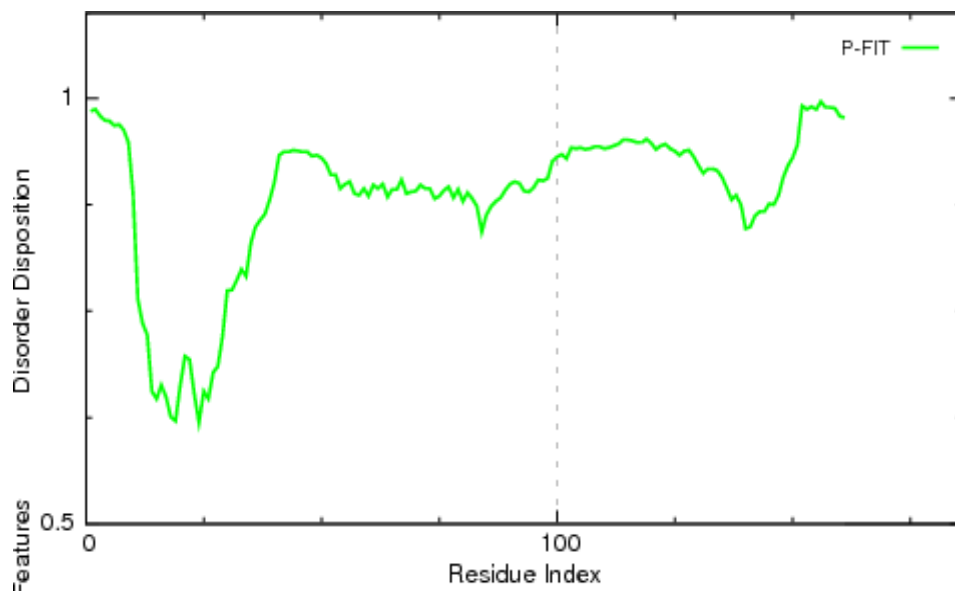
Amino acid bias of intrinsically disordered regions:⁴

Low content: C, V, L, I, M, Y, F and W

High content: Q, S, P, E, K, G and A

LQIAHLHRPGLGSAGPEPAFARDHLRDFGRAPSDRPPAMPGTAPAPAAVPVSAATEGAPVSAP
TGVPARAAVRPAVPAPTAVPTLTLRGAAEPARRPATDPMKPPMTSDRKISLRPLAPATAAT
RPAAPATPTPAAPAAVPAPPAAPAPAAVPAGRQE

b



C

```

BryC_Module11_ACP      1 ENYDRKKMMDHIVNEGCTMVSIDDEL---IEST---Q---SLNTDATSFIPDLNLFIEE-----KNETDIYHIESYD-----DSYDDL
PksM_Module11_ACP      1 -----ISSGTQQSEAV-----KQ-----HSQDMK
BryD_Module13_ACP      1 -----ISEQYGDVACQEDKVNIN---VNSEK-----SISQNEI
PedH_Module14_ACP      1 -----LGGNAGRVLVLA-----
BryX_Module16_ACP      1 -----IVKDRSTHT-----HCD-----VMTSTKAISSK-QLPLGKNKEILKNNDSFVGGDDV
PksL_Module6_ACP-ACP   1 -----IDVNAKAE---KRT-----EPPFAPVQPVIPKPSVDREASGKPANIT-LQPLMTN-----QD-
BaeL_Module6_ACP-ACP   1 -----IDTSA-DAAVKRAETQSAPIAPAAAPEPVIEVTLHEETVKE---KPSNIT-LQTLGSAKPKRTVQ-----ERV
OzmQ_Module1_ACP       1 -----VEPEHRT-----ARRTPEPDTL-TDRLGAAVGH-----ERV
OnnI_Module4_ACP       1 -----FSSQTQDVDLK-SQSIETPVTETAPQP-VLQQDNV
DifL_Module14_ACP-ACP  1 -----AADVNSSLPGQAMKSETQTVLPSSPQPEHEGLISLSPLTEIEMTAFEPETLQKPIV-LQLDCTVSVQQPEK-AVSEDNR
PedF_Module4_ACP       1 -----
VirH_Module9_ACP       1 -----
ChiC_Module9_ACP       1 -----
MmpA_Module7_ACP-ACP   1 -----APSEVAVSVE-----
MmpA_Module5_ACP       1 -----
LnmI_Module2_ACP       1 -----LAAPA-----
DszC_Module11_ACP     1 -----L
AlbI_Module2_ACP       1 -----L
ChiF_Module19_ACP     1 -----IGQGEAPFWSQPSQ-----QT-----RRNGEAEHAEPPP-RL-----
LnmJ_Module5_ACP       1 -----IGG---WRS-----AQETGESREPVETA-----
EtnE_Module6_ACP-ACP   1 -----
EtnD_Module1_ACP-ACP-ACP 1 -----
KirAII_Module5_ACP     1 -----ITPGP-RFDQAQA
OnnB_Module3_ACP       1 -----VARRKETGMLTEQNHEQREPTSEKADHEFFWKIEELIEE---WER-----QQYGGHLTPETLTK-VFAYEVV
LnmJ_Module8_ACP-ACP   1 -----
ChiD_Module11_ACP     1 -----V
DszB_Module8_ACP       1 -----
SnaE2_Module5_ACP-ACP  1 -----YDLQIEHL-----HH-----KGLG---TATAEP-ALSRPHL
DszC_Module10_ACP     1 -----
KirAIV_Module9_ACP    1 -----
OzmJ_Module10_ACP     1 -----
VirA_Module5_ACP-ACP   1 -----YDLQIAHL-----HR-----QGLG---SAGQEP-AFARDHL
EtnI_Module20_ACP     1 -----
consensus              1 -----

BryC_Module11_ACP      71 KDNDNMSEM---SDDI-----
PksM_Module11_ACP      20 TEIDEPNGKT-----
BryD_Module13_ACP      31 TQQ-----
PedH_Module14_ACP      13 -----
BryX_Module16_ACP      47 AETTLMVD-----
PksL_Module6_ACP-ACP   50 -RLERVPSDTETETIT---AE-----
BaeL_Module6_ACP-ACP   63 -----PASRPAVSAE-----
OzmQ_Module1_ACP       31 D-----
OnnI_Module4_ACP       34 LEKKPLPETRSTETS-----SCGSPREQQV-----
DifL_Module14_ACP-ACP  80 -----
PedF_Module4_ACP       1 -----
VirH_Module9_ACP       1 -----LPVPAASARP-----EKEQQPD-----
ChiC_Module9_ACP       1 -----ITQAPVSAPAPEAPPVTD-----ANGTA-----ASPSTLRAPQRGTSLSL
MmpA_Module7_ACP-ACP   11 -----SVASVAQGGKIRL
MmpA_Module5_ACP       1 -----
LnmI_Module2_ACP       6 -----ARR---P-----DGAAGSAP-----AAPESGQSAPP-----ASP--QVQDDRAD--
DszC_Module11_ACP     2 DDEALPHGAWSATAAPPAQTAWSATAAPPARAA-----
AlbI_Module2_ACP       2 RDD-----AAHAAEPAP---ADG-----AAEDASADAPNAANAPTDPV-----ATL
ChiF_Module19_ACP     34 -----
LnmJ_Module5_ACP       21 -----HAPTAVPAAPV---DQADLPEVREERFGLDPQ-----
EtnE_Module6_ACP-ACP   1 -----GPWPAPSAAPPATAPSRPASPPTSIALAGA-----DAAVRSDAASPPAIARSAPKERAERIRADAAPISL
EtnD_Module1_ACP-ACP-ACP 1 -----
KirAII_Module5_ACP     13 EAAGLA-----
OnnB_Module3_ACP       65 KNLGPEQIHR---VDQLL--AATPKPATAPPPVQAAA-----VVPEPKLVRSED-----
LnmJ_Module8_ACP-ACP   1 -----FPTDAEP-----TAPAVPARKEA-----TMTARS-----EAQ--PTRRGPKIRL
ChiD_Module11_ACP     2 REISRASGAA---PAPAAAAPAPALAAAPAVAA--AFAAAPAPA--P-----
DszB_Module8_ACP       1 -----
SnaE2_Module5_ACP-ACP  28 RDFGRPHTPA---P--AAAPAPSPAADPAPTAPRRAPEAAPAPAPVPRRVEPAPALVPRPAPASAPAPGAVTPATDRDPQETP--FMTDRKISL
DszC_Module10_ACP     1 -----LGRPAGDAAAP-AVARGETAEEAPSRGET-----AEEAP--SRG-----
KirAIV_Module9_ACP    1 -----KLFTQPAAR---P--VAQPAA-----EPAPAPRTEPVAEP--LAGAGAG--
OzmJ_Module10_ACP     28 RDFGRAPSDR---P--QAMPGTAPAPAAVPSAAEGAQVS--APTGVPARAAVVRP-AVPAPTAVPPLTLRGAAPARRPATDPMEKP--QMTSDRKISL
VirA_Module5_ACP-ACP   1 -----
EtnI_Module20_ACP     1 -----
consensus              201 -----

```

```

BryC_Module11_ACP -----
PksM_Module11_ACP -----
BryD_Module13_ACP -----
PedH_Module14_ACP 13 -----SEHS
BryX_Module16_ACP -----
PksL_Module6_ACP-ACP -----
BaeL_Module6_ACP-ACP -----
OzmQ_Module1_ACP -----
OnnI_Module4_ACP 59 -----LAE-----
DifL_Module14_ACP-ACP -----
PedF_Module4_ACP 1 -----VDSAVVEAERAPN-----SKAPASMLS
VirH_Module9_ACP 19 -----GPT-----GTGPT-----PF-----
ChiC_Module9_ACP 40 D-----
MmpA_Module7_ACP-ACP 24 -----RALDITLPVADSAPIVPPAPRILSQPAPLAAKAAFPV-----PVADD-----
MmpA_Module5_ACP 1 -----YDAPAPMAAYPAPGTSRT-----GRRAWSPD-----
LnmI_Module2_ACP -----
DszC_Module11_ACP 37 -----DPGGA-----APPEG-----PGGAPPGGGAARQ-----
AlbI_Module2_ACP 40 VRRITVAQVLGYF-----
ChiF_Module19_ACP 34 -----
LnmJ_Module5_ACP -----
EtnE_Module6_ACP-ACP 65 VALVPPAAMDTAASSAAR-----SAP-----VAISLSPLRAEPASAAY-ARRADEPPASGGE-----
EtnD_Module1_ACP-ACP-ACP 1 VTR-----RQVAPAPEVSPPTVAP-PGPAAIKPKSGIALR-APGEERFAAPA--VSSGVEHTVLSLSPVGAAVNQAAVAVGSAAGAAAPYGGGSPAGGAASSAIA
KirAII_Module5_ACP 19 -----GTAQDEVTTWPEP-----AAAVP-----APPAESAPAPAATAAGGDPES-----
OnnB_Module3_ACP -----
LnmJ_Module8_ACP-ACP 39 AAP-----RSAGGASATAPTREPVPAP-----PAPAPAP-----A-TTAPQVVRADVPPALPEAAAAA-----
ChiD_Module11_ACP 43 -----ALAAPAVAA-----AFAPAA-----PAPVTPAGSQ-----AEDGLLD-----
DszB_Module8_ACP 1 -----LDPAPS-----DEASPSFA-----PPPEAPR-----PAAAP-PAPPSAEAR-----
SnaE2_Module5_ACP-ACP 120 -----RPLGTAPTAPAPAPAPAPVSVAPAPV-----PATPT-----PAAAP-AAPVAAPSVAAPVAGAVD-----PAVA
DszC_Module10_ACP 37 -----ETAEEAPSRE-TAE-----EAPALA-----PATADP-----ALRK-----
KirAIV_Module9_ACP 2 -----RSAAPRAATPTPTRP-AGSSGP-----
OzmJ_Module10_ACP 38 -----AVAEAAAAGAEFY-----APERLA-----A-----
VirA_Module5_ACP-ACP 117 -----RPLAPATAATRPAAP-ATPATPAAPAAV-----APPAAPA-----PAAVP-AGRQE-----
EtnI_Module20_ACP -----
consensus 201 -----

```

```

BryC_Module11_ACP -----
PksM_Module11_ACP -----
BryD_Module13_ACP -----
PedH_Module14_ACP 17 ERDAVEFEVERNGSA-----
BryX_Module16_ACP -----
PksL_Module6_ACP-ACP -----
BaeL_Module6_ACP-ACP -----
OzmQ_Module1_ACP -----
OnnI_Module4_ACP -----
DifL_Module14_ACP-ACP -----
PedF_Module4_ACP 23 G-----ER-----
VirH_Module9_ACP -----
ChiC_Module9_ACP -----
MmpA_Module7_ACP-ACP -----
MmpA_Module5_ACP -----
LnmI_Module2_ACP -----
DszC_Module11_ACP -----
AlbI_Module2_ACP -----
ChiF_Module19_ACP 34 -----NGTINGAHAD-----
LnmJ_Module5_ACP -----
EtnE_Module6_ACP-ACP -----
EtnD_Module1_ACP-ACP-ACP 94 GVERVTPRATEGGEGGHMPGTTATARDRRASPVDNAAGGRSLA
KirAII_Module5_ACP -----
OnnB_Module3_ACP -----
LnmJ_Module8_ACP-ACP -----
ChiD_Module11_ACP -----
DszB_Module8_ACP -----
SnaE2_Module5_ACP-ACP 179 D-----
DszC_Module10_ACP -----
KirAIV_Module9_ACP -----
OzmJ_Module10_ACP -----
VirA_Module5_ACP-ACP -----
EtnI_Module20_ACP -----
consensus 301 -----

```

PKS Module	Linker length
VirA_Module 5 (KS-ACP-ACP)	165
SnaE2_Module 5 (KS-ACP-ACP)	179
PksM_Module 11 (KS-ACP)	28
EtnI_Module 20 (KS-ACP)	10
OnnB_Module 3 (KS-ACP)	110
DifL_Module 14 (KS-ACP-ACP)	79
OzmJ_Module 10 (KS-ACP)	56
PedF_Module 4 (KS-ACP)	25
OnnI_Module 4 (KS-ACP)	61
MmpA_Module 7 (KS-ACP-ACP)	64
MmpA_Module 5 (KS-ACP)	26
OzmQ_Module 1 (KS-ACP)	31
KirAll_Module 5 (KS-ACP)	58
LnMJ_Module 5 (KS-ACP-ACP)	91
EtnD_Module 1 (KS-ACP-ACP-ACP)	136
BaeL_Module 6 (KS-ACP-ACP)	72
PksL_Module 6 (KS-ACP-ACP)	66
BryX_Module 16 (KS-ACP)	54
DszB_Module 8 (KS-ACP)	36
EntE_Module 6 (KS-ACP-ACP)	116
LnMJ_Module 8 (KS-ACP-ACP)	49
ChiF_Module 19 (KS-ACP)	42
KirAIV_Module 9 (KS-ACP)	22
AlbI_Module 2 (KS-ACP)	51
DszC_Module 10 (KS-ACP)	65
ChiC_Module 9 (KS-ACP)	40
BryC_Module 11 (KS-ACP)	84
BryD_Module 13 (KS-ACP)	33
PedH_Module 14 (KS-ACP)	31
Lnml_Module 2 (KS-ACP)	39
ChiD_Module 11 (KS-ACP)	57
DszC_Module 11 (KS-ACP)	98
VirH_Module 9 (KS-ACP)	28

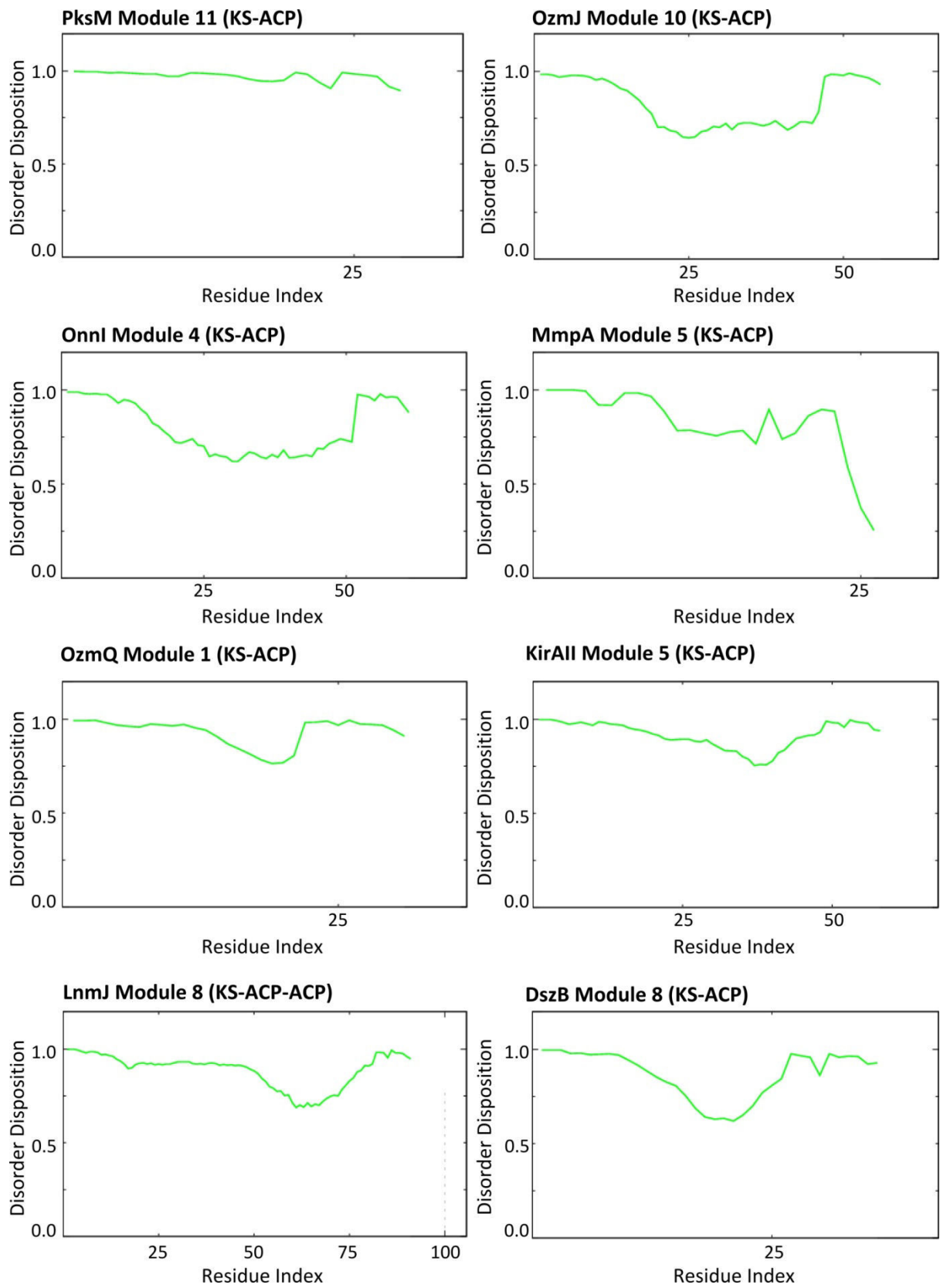
d

```

VirA_Module5_ACP-ACP      1 YDLQIEHLHSCGLGSAQQBPAPARQHLRDFGRAPDRICOMPETAPAPAAVVSAREGAQVS--AFTGVFARRAVRPA-VVEAFVAVPTLTLRGAAPFARR
SnaE2_Module5_ACP-ACP    1 YDLQIEHLHSCGLGSAQAPALSRQHLRDFGRPHPAARAPPSPADPAETPAPRRRPEAPAPAPVRRVVEPAPALVEREAPASAPAAPCAVTEPAT
consensus                 1 ***** ** . ** . * ** * * * * * * . * * * . * * * * * * . . * * * * * * * * * * * * * * * * * * * * * * *
VirA_Module5_ACP-ACP     98 PATDEMEKQEMTSDRKISLRPLTPETAAATRPAPAPATPAAPAAVPAAPAPAPAAVPAER-----C-----
SnaE2_Module5_ACP-ACP   101 DRPDQETPQPMRDRKISLRPLTPETTPAPAPAPAPAVSVAAPAVPAT---TPAAAPAPVVAAPSVAPVAGAVPAVAD
consensus                 101 ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

e



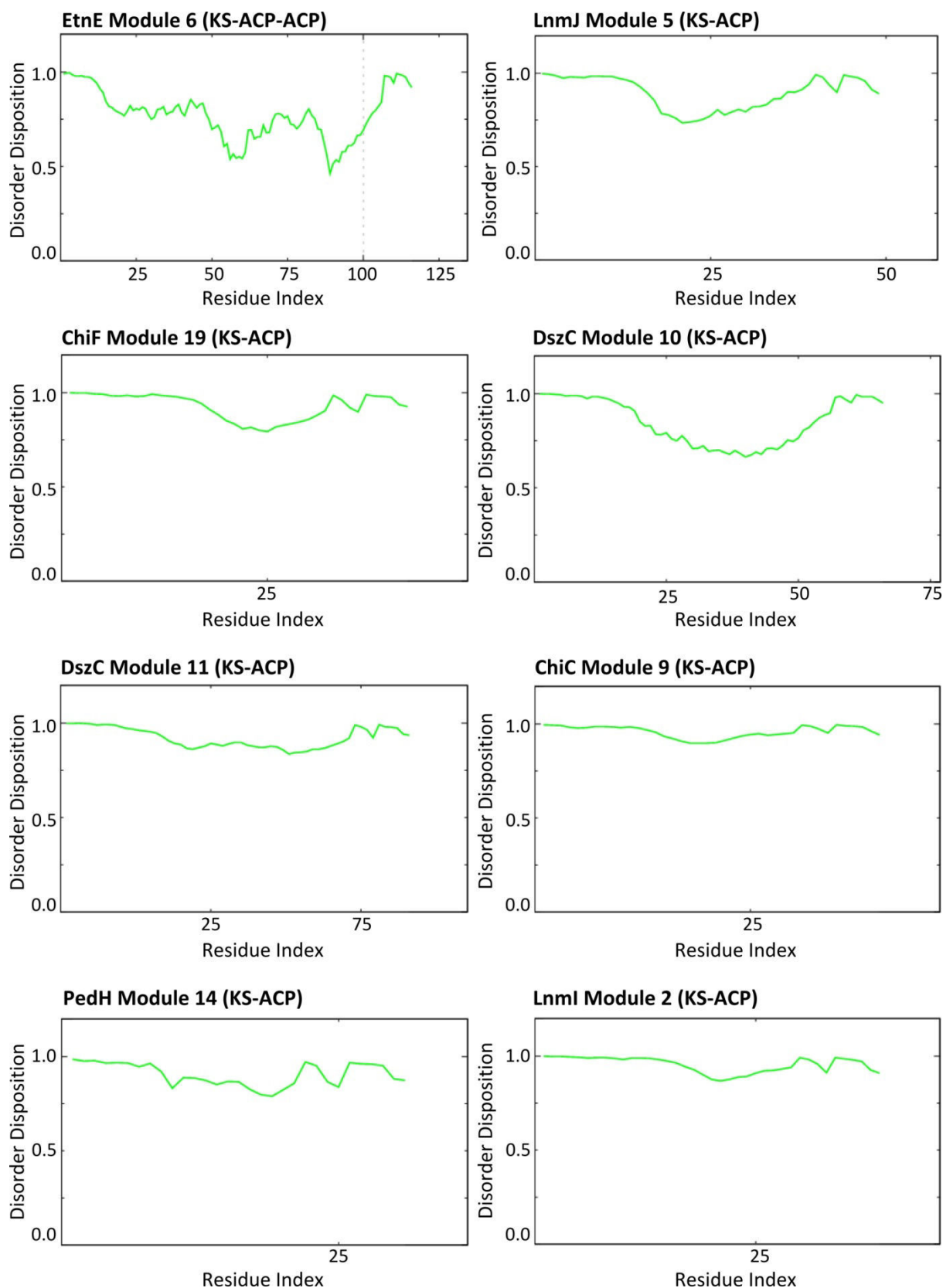
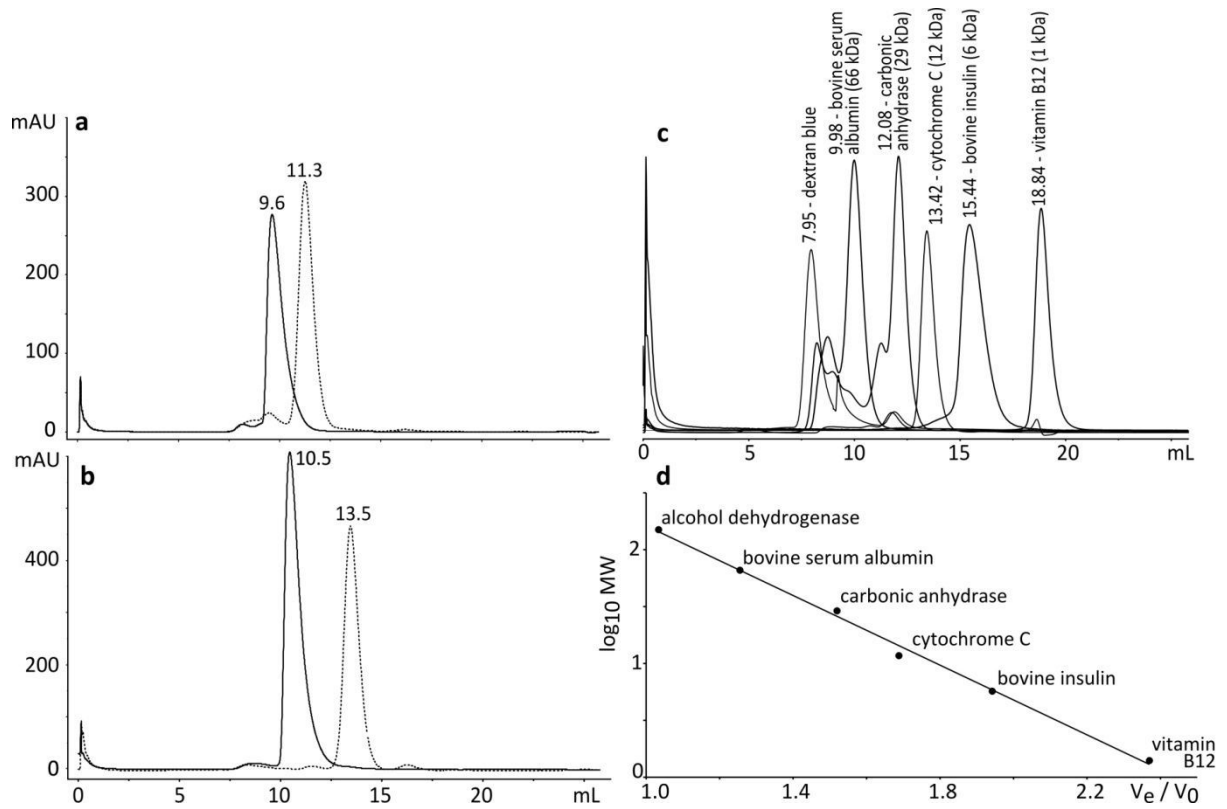
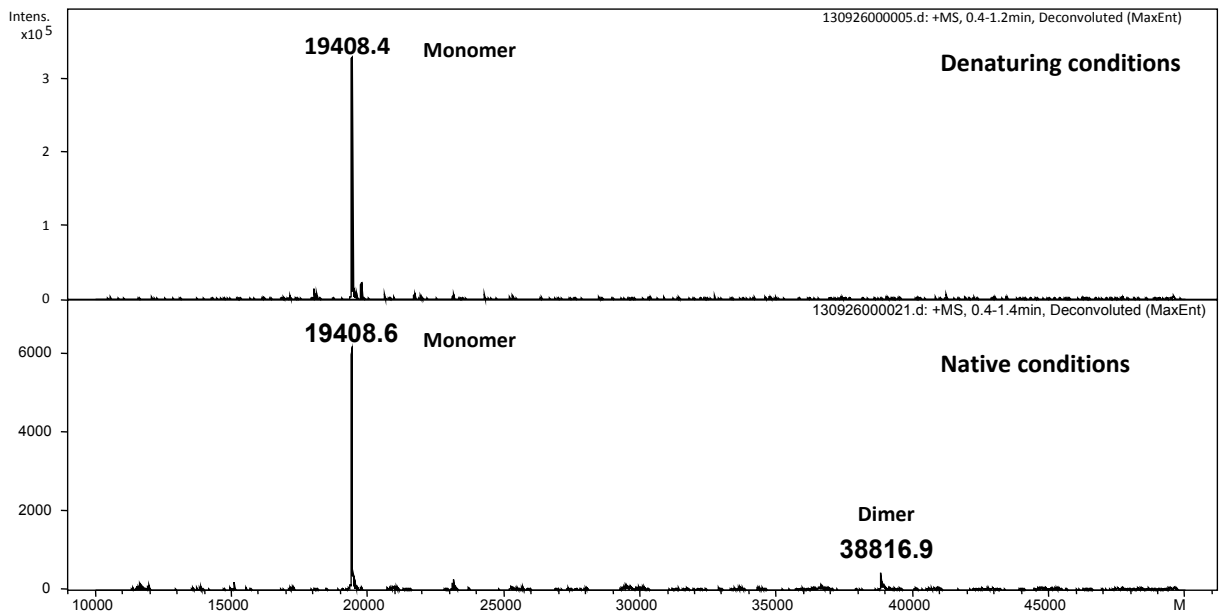


Fig. S6 Sequence analysis of the linker region following the ‘post-AT Linker’ (referred to as the ‘post-post-AT-linker’ (Fig. 3)) in various *trans*-AT PKS. (a) Analysis of the amino acid composition of the 159 residue post-post-AT linker of VirA module 5 reveals a sequence bias consistent with an intrinsically disorder region.⁴ (b) Disorder propensity analysis by PONDR-FIT⁵ of the post-post-AT linker from VirA module 5 (<http://www.disprot.org/metapredictor.php>). (c) Alignment of the post-post-AT linker

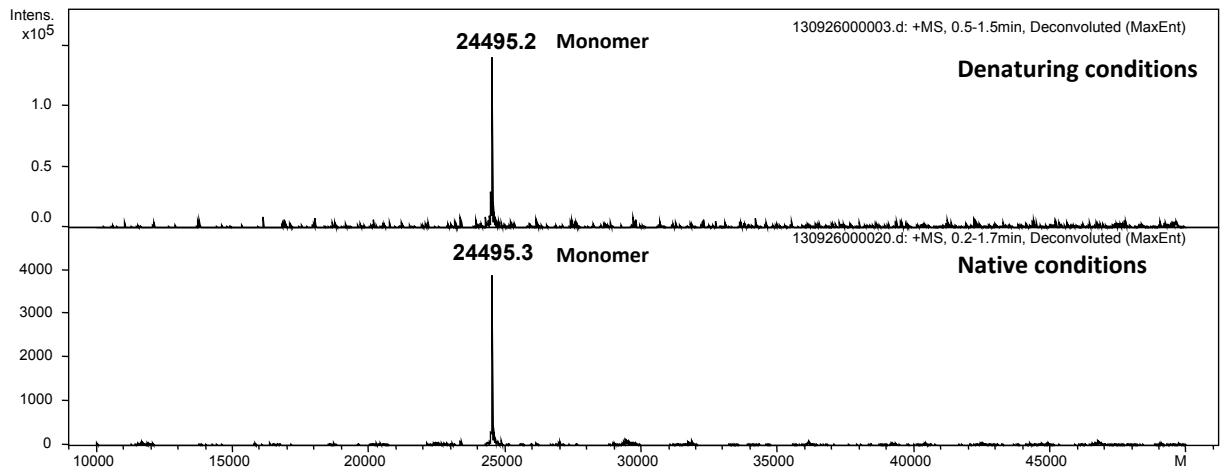
from a selection of KS-(ACP)_n modules from *trans*-AT PKS. The subunit name, module number and its ACP composition are indicated. The linker lengths are given in the table, with the two extremes highlighted in pink. (d) Comparison of the linker regions following the post-AT linker in virginiamycin and pristinamycin modules 5. (e) PONDR-FIT analysis of the linker regions in (c) which show the highest disorder propensity. Key: Vir, virginiamycin; Sna, pristinamycin; Pks, bacillaene from *Bacillus subtilis*; Etn, etnangien; Onn, onnamide; Dif, difficidin; Ozm, oxaxolomycin; Ped, pederin; Mmp, mupirocin; Kir, kirromycin; Lnm, leinamycin; Bae, bacillaene from *B. amyloliquefaciens*; Bry, bryostatin; Dsz, disorazol; chi, chivosazol; Alb, albicidin.



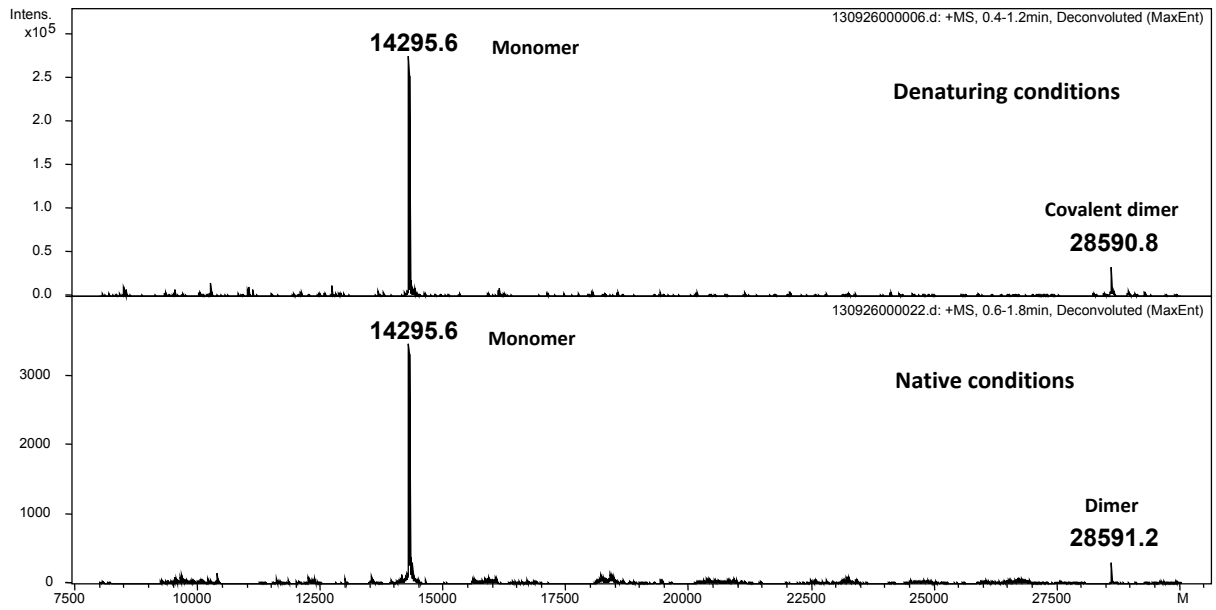
e



f



g



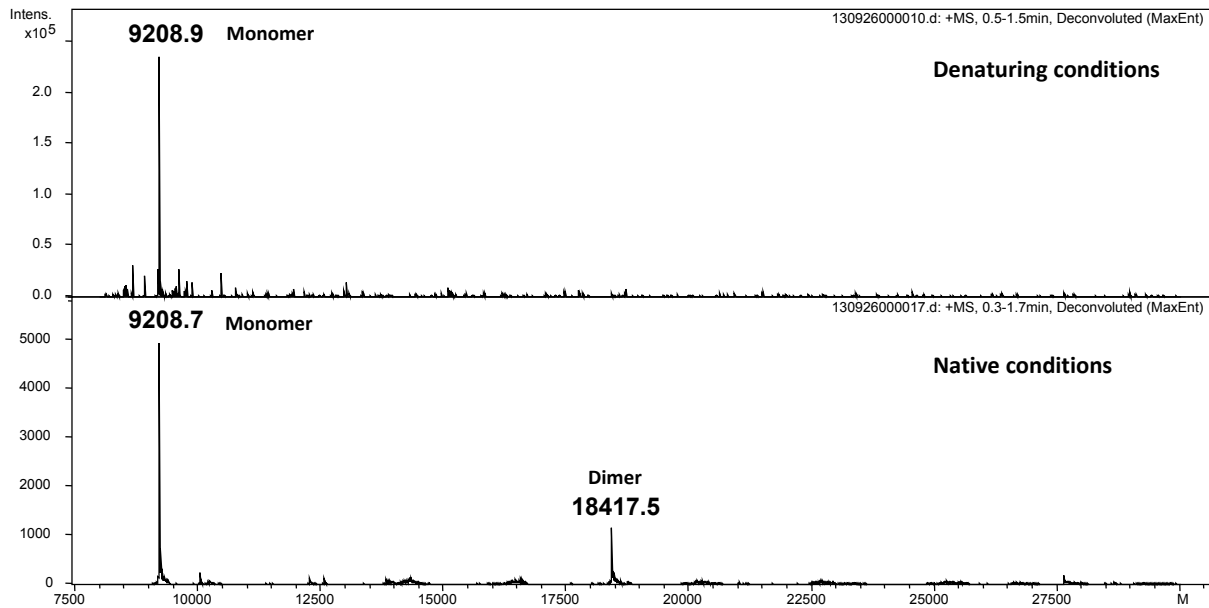
h

Fig. S7 Analysis of the effect on the oligomerization state of various constructs of the presence of the putative dimerization domain (DD). FPLC traces of (a) ACP_{5a}-ACP_{5b}-DD (solid line) and ACP_{5a}-ACP_{5b} (dotted line), (b) ACP_{5b}-DD (solid line) and ACP_{5b} (dotted line), and (c) the calibration standards (Sigma) dextran blue, bovine serum albumin, carbonic anhydrase, cytochrome C, bovine insulin, vitamin B12 (alcohol dehydrogenase omitted for clarity). All samples were analyzed on a Superdex 75 10/30 column (GE) equilibrated with 50 mM sodium phosphate (pH 7.5), 250 mM NaCl. (d) The calibration curve plotting the ratio of elution volume (V_e) to column void volume (V_0) against log molecular weight. The curve was described by the formula $y = -1.53x + 3.74$, with $R^2 = 0.996$. The calculated molecular weights are shown in the **Table** below. (e) Analysis by mass spectrometry of ACP_{5a}-ACP_{5b} under native and denaturing conditions shows that the construct is almost exclusively monomeric, with a small proportion of dimer (estimated at 5%) present. (f) Analysis by mass spectrometry of ACP_{5a}-ACP_{5b}-DD shows that the construct is exclusively monomeric under all conditions. (g) Analysis by mass spectrometry of ACP_{5b} shows an estimated 18% of dimer under native conditions. (h) Analysis by mass spectrometry of ACP_{5b}-DD reveals a small proportion (estimated at 9–11%) of dimer under both native and denaturing conditions, suggesting the presence of a covalent link between the monomers (disulfide bridge?). Panels e–h show the deconvoluted neutral mass spectra obtained using maximum entropy analysis.

Construct	Expected MW / kDa	Estimated MW / kDa	Ratio of MW in presence/absence of the DD
ACP _{5b}	9.2	14.1	
ACP _{5b} -DD	14.3	53.5	3.8
ACP _{5a} -ACP _{5b}	19.4	37.4	
ACP _{5a} -ACP _{5b} -DD	24.5	77.0	2.1

The calculated molecular weights of monomeric constructs with and without the putative dimerization domain, compared to the estimated molecular weights derived from calibrated gel filtration (**Fig. S7**).

Commentary on the table

The presence of the dimerization domain (DD) on ACP_{5b} gives a molecular weight approximately four-fold greater than the ACP_{5b} alone, while that of the ACP_{5a}-ACP_{5b} is more than doubled by addition of the DD. It could be argued that the addition of the DD makes the proteins substantially more elongated, thus explaining the apparent increases in molecular weight; indeed, an elongated protein can elute at twice the volume of a

globular protein of the same molecular weight.⁶ However, the fact that the ACPs even in the absence of the DD give estimated molecular weights in great excess of their calculated weights, shows that the constructs are already in an extended conformation. Thus, our analysis instead supports a change in oligomerization state in the presence of the putative dimerization element.

MDAKEILTRFKDGGDLDRAAAQALLAGRTPAAAPRPSEAAAPTRPAVPAVPAVEPAAGTTVAEAGTAGRPEPVA
VIGYSARFPGAADADTFWQRILDGDDLVEVPPERWRTEEFYDPDPAEGRSVSRWGAYVADADRFDADFFRMT
REAELTDPQARLFLQEAWRALAHAGRDARSLAGTRCGVYAGVMLNDYQDLVERESPYKRLPQVMQGNNSILAA
RIAYHLDLKGPAVTVDTACSSSLTALHLACQSLWLGETDLTVVGGVTLYLTELPHVFMSAAGMLSPNGRCRPFDA
ADGIVPGEGCAVVVLKPLSKALADGDPVHAVIRASGLNQDGRNTGITAPSARSQTALVRDTLQRFVDPAGIDYVE
CHGTGTPLGDPPIEVTALNEAFAGAGLAPASVPIGSVKNIGHTSAAAGLAGLLKAAGVVRTGLVPPSLHYARANPQI
PFDQGPFTVAGERRELGRPDGGRPRRATVSSFGFSGTNAYVVVEQAPEQAARPAGDDGAPPVLPVLSGRRADAPAA
HAADLARWLRGPGSDASPADVAHTLAVARTHHTYRFALLVSGRDELLESLDLLAAGSADPRRTDTAPDAAAPDAA
VRRQAALLARLVERAGENPGPVELAALAKLYTQGHVTDWAAVSPPARHRRLSLPAYPFAPHRHYVERPAPVTPP
TPAATSALAEASHGTPAEPAGARTPQEALADVPLTEPVLHRQEWIPAPLPAAPAPTAPVLVHDPAGDLAAALAAAT
GLDVHRLGGDRTAEEALRATGAQVVTVVLRLTPPADGAPATAVDLAAFEAARAALAVLRTQQLTLLAVTADPTR
ADAAGALVQTLRQENPRLSGRAVLTADGTPDARRLLAEIAAPAAEHGHLADLRSARRLRRVLVPVPLPGTRPFVRQ
DGVYLVSGGAGGIGLALARHLTDRPGTRVVLVCGRTAWDDLAEATRSVSGSDRLRYARADVTDPEAAAAALVAEV
VRTEGALHGVFHAAGVVRDGYLVRKDPDADVLPKILGARALDEATAALPLDAFVLFSSVAAVTGNLQSDY
AFANGFLDGFATRRAGQVARGERHGRTLSVQWPLWDVPGMSIPEPVLEVVAQHTGMAPLPAAVGLAALERLLAA
DGPEVVSFLFHGDAATWRAHLAALHLERPAAGQVTAAPAAPAPSDTPVVASADPAAADDGRRAELADRVSRTV
ADTIGRPAGSIGGHTSLESMGLDSVMIRALASRLSAEVAPVGPPEMLFGLRDLDELVDHLVAARPVPVAEPATPAVPP
APAATAATVFAPVAPADPAVPPAAPAALPVPAAPASASRTLPAPSADDRFAIIGISGRYPQAPDLDAFWQNLNGK
DTASDLPTDRWPDAAGVNRGHFLEGVDADFPTFFGLSAHDGTLDDPQERLFLEVAWEALEDAGYTGSRHLDLVA
PDGERRSVGVFAGITSSDYKLLGAERWAAGHREMPGSHYWSLPNRLSYLLDLRGPSPQVDTACSSSLVALHLALDA
LRRGECAAALVGGVNLVYHPSRFRMLRRSGFLAEDGLCRSFGAGGAGFGPGEGGAVVVKPLAVALADGDTVHA
VVRGSAVAHGGRNGFTAPSPWAQARVLRALRNSGTDPETVNIEAHGTGTELGPVELAGLQDAYGSGRVPCS
LGSVKSAVGHGESAAGIAALTKVVLQLRHGELVPTLHADPVNPLRLLEDTRFVLQHTPGRWERLTGADGSPLPRA
GISSFGAGGVNAHVIVEEYLPEPHGRAAAAEPGRPELVLLSAPTREHLAATADRLARRLEGPEAPADLRAVAYSRT
GRSAMDCLAVVATDTAGLAAALGAFARSADAGEGESAVRYADLRDGRRAHRDLDTVPETTAFLADLWRNRRHL
QLGELWLSGLDIERAAPREPGRTRVPLPPTSFLRRRLWITDPVTADTTPTPDPLTPDPLPVPLSAPAPAPAPAPQAP
TTTPAPAAAPAAAPARSGEAVTEQLSRLVAGFVPDAAGSVDPDRITLLEHGIDSINLMNLRFEITERFGRTLPLQLLSE
STVPVLAHLSADRAHDRA*

Fig. S8 Corrected sequence of VirFG. The predicted natively unstructured region is shown in bold.

a

```
Ta-1_ACP-KS      1 -----TIVEIWKNVLQV---NEVGVDFRFFVGGDSVLAAVLVEEMNRRFDTRLAVTDLFKVYVNTIR
BonA_ACP-KS_2   1 -----AEEMQIPR-EAISAHTNLAEYGFDSIALTEFARRLARHFSHELEPTLFYSHPSLIG
Bat2_ACP-KS     1 -----QLKEQARELLKLPY-DRIGQEVNLVDFGSDIALVTFAKRLSQCFDFVLPSPVFFSHSTLIG
VirA_ACP5b      1 -----VVARALREELARLYCEP-CDID-EASFNILGLDSILGVE-VAEVNQTGYGLDEKAGLLYD-PSLA
OocJ_ACP-KS     1 -----LAKK-VDIDETKAFSEYGLDSFAAINLAVLNAAFSIKLEPCTVLFDDYGCVS
Ta-1_ACP-KS_M   1 -----IREELATSIAQALYIDR-AQVNAESTFVEIGLDSIVGVEWTHAINNQYGLSLPVTVVYDHPNLS
EtnE_ACP-KS_M   1 -----APRPVEELVAGLTQTLAAALYAA-SDIDEDRLFVHGVDVIVSVEWVRINATFHIDLRATLFLFDHPTIH
BaeL_ACP-KS_M   1 -----HDELSKALADVLYMER-HEVDIDEAFILGMDSIIGLEWTKAVNKRYCTDCNVTKVYDYPTIR
OocL_ACP-KS     1 -----DEIQNDRAFAALGLDSILGAEWTHLNLNLTIGTELATRLYDYPTIQ
BonA_ACP-KS_M   1 -----IDVDRFVALGLDSILGVEWMAINRRRHGLALNATLVYFHPVTR
BonA_ACP-KS     1 -----RAVIRRESIGQALKIGE-SQLDDEAFSNYGLDSITGVAVETINARIGLIDLEPTVLFDDYVSE
PedI_ACP-KS_M   1 -----VEVGLDSTAVETWIKALINQRYGLSITGATRVVSWSCIA
OnnB_ACP-KS_M   1 -----DVFVEMGLDYVMAGSWVQKLNQAYGLSLBATVIYVYTNLL
PsyA_ACP-KS_M   1 -----YKILRETLARELHMV-EDIDDRPFLDMGLDSVIGVTVWRKLNREHGLSITVTKVYAHETVC
EtnD_ACP-KS_M   1 PRDARVGAARSSELDALQVEIAKSLADALYMEP-CDVDVDRPFVILGLDSIVGVEWTHAINNAYGWSLEPATRLYDHSVTR
Bat2_ACP-KS_M   1 -----ARSPAMSHEELSAELVQGLAKVLYLEH-DDIDRDTAFSELGLDSITGVWVRLNRRYQTSLEGASRLYDYSTIK
Bat3_ACP-KS_M   1 -----ADLRMEIAESLAKVLYLEN-DDVDMDTAFSELGLDSVIGVWIRBELNQRYGTSLEGASQLYDYPTIK
OocN_ACP-KS     1 -----IGRILGNGADDETCQQAFTELGLDSVVAAFVSELVAAYVWKRKRSDFVFSHTTPA
PedF_ACP-KS_2   1 -----ARRQDVAARVRELLDLSLAQALSIGR-EQIEQDIFPSDYGLDSILGVGFVQLNDELGLSLNNTLLFDYTTVQ
PsyD_ACP-KS_2   1 -----VPQPASEDQIASHVVDALTRVLRMEN-GRIDPHVAFSDYGLDSILGAEFIDRVNNAALGLALNNTLLEFETTTA
OnnI_ACP-KS_2   1 -----LEVDPHGELVHRLVSHLAKVLDIAE-STIEGDVFPFSDYGLDSILGVNFIQTQINDDLGLEMNTTIVFCHTSVN
BonD_ACP-KS_M   1 -----AIDATPEFILLGLDSITGVWVRKINRHGLSLIAMQMYEHPETLR
EtnE_ACP-KS     1 -----REEALRQVKKILAPVTKIPV-ERIDAAASFETYGLISVMVAVELTDLLEA-VFGPLSKTLLFEAKTVR
EtnF_ACP-KS     1 -----AREGASLLHATLARLELLAVVTTIEA-ERIDGAEFPPEAYGLDSIMITQINQLEA-IFGALSKTLLFEHRTIE
EtnD_ACP-KS     1 -----HADVTKIPF-SQIDPEESLELYGLDILLIRLNEALS-EVFGDISKTLLWFYQVTA
PedI_ACP-KS     1 -----LKKIVARILRMS-GELAVREPLETYGLDSILIVQNTNALRE-VFPEIPSTLLFEQCTID
PsyD_ACP-KS     1 -----DQLRRLEADVMRISV-DDVIVQAPLESEGLDSVVVVTYLNQRLRD-AFGEVSAITLFEYRTIA
OnnB_ACP-KS     1 -----YFQTLIGKVLRIE-ORIDPAESLEAYGLDSILIVQIIEALQE-IFEDVSSITLFEFVQTV
PedF_ACP-KS     1 -----QLKRIILAQVIGRAV-ERISCEPMDRYGLDSIAITQINRLEE-QFGGLSKTLLFEYQVTE
OnnI_ACP-KS     1 -----AEIDDQELATRTLQEMKTLGSGVIG-DEIDAQKPLENYGLDSIAITQINRLELDG-VFADLSKTLLFEYQTV
VirA_KS         1 -----
BaeM_ACP-KS     1 -----LKKILADKLGQFW-ETHVLAGVYILGLDSSSLEILVQDLSKKTICADLAPTLFFEYTNLK
Bat3_ACP-KS     1 -----EQGLIADKLGAP-EQIDRDSGLYELGWSAGLLELVDDIEQKIDASLPPTLFFEYVTTG
OocL_ACP-KS_M   1 -----EVLAIISVILKIPV-EKIDSEKISYGVDSIITEHMGRIAGALGWSLSAAIFFEARNIE
PsyD_ACP-KS_3   1 -----QTTAADGLLKEVGAHVELVAVVQKIQP-EKIRLNRELAHYGFDSISFTTLANALNAYDLSLMPTEFFETENEA
Bat3_ACP-KS_2   1 -----ALMQALSEBOLKVKV-ENVVDVAEISEEFGDSISLTTIGNLNLNRYGLELSPTEFFEYSTLR
Ta-1_ACP-KS_2   1 -----AVLLEGVSEELKVP-EEEDADTKLSYGFDSITFTTEFANLNLNQLSGLSPVFFFEETIAN
TaO_ACP-KS     1 -----QVLMVSVSKVLPF-EEEDADAELKMGHDSISLTLALQOLAGEVRLDLAPTEFFEETIN
BaeJ_ACP-KS     1 -----DADKSLQTHHTALTAMVSGHLKVR-EDIEETDTELSHYGDSVSFTVFTNQLNAYQLELAPTEFFEAGTIS
BaeN_ACP-KS     1 -----DADSLLEKVKHLRQQTASLKVNI-EKIDPHEEMRYGLDSISMTFTNQLNRYRLLELPTLFFEHPHTIH
```

```

Ta-1_ACP-KS      59  DVARHVECATQARTGATE-----
BonA_ACP-KS_2   55  QFAAFLLASHGERLAAFYDAGGVAAALS-----TAPASRTS-----VALPDEPRVDV-GAAAPLV
Bat2_ACP-KS     61  KLTSHFVVEHADLRYRQQEHVVGART-----RVTVRGE-----VSMP---AVDIPGLAREAV
VirA_ACP5b      65  ALSHVAG-----
OocJ_ACP-KS    51  DLRRYLLENYRHEVARMQPAQRQDAASA-----GNAAPRAE-----T
Ta-1_ACP-KS_M   64  LFAEYLARQLPTAPPPAPP-----QVEAKT-----VRVAPVVSSEPP-----S
EtnE_ACP-KS_M   71  QLAHHLATLDATAVRAAVDGISSAERAGAHESTARREREAAEREAARRAREDAERQVAARHEPEASFEVEARRASDA
BaeL_ACP-KS_M   63  QFAEFLRQPSVVRGKKQ--VP-----VR-----PKPLTQOHAP-----QEKTPAE--ERLTI-----
OocL_ACP-KS     47  VLAEYLTQSHGATPR-----PKPNAAVSEPEA-----PC-SQDL-----K
BonA_ACP-KS_M   45  RMAARLALALAEATFAAAMPAPATESAPAPA-----PAPAPASVPAQA-----PAVARDLPAA--GPLPAAAPASATDH
BonA_ACP-KS     63  QLSREAGAYASRFRVPGTEPMPMPRSFVP-----P-----E---AAAAPAVRV-----A
PedI_ACP-KS_M   38  EFSQYVRSLLAAQFAAVSRAPAESDRAESLLLRRFEPDVTNQPRE-----EQEE-ASTQDEPIVRL-----L
OnnB_ACP-KS_M   42  DLAGEHLASEMSHRLATA-RPLAKLERERPRRT-----P-----RADA-----KPVVS-----P
PsyA_ACP-KS_M   64  AMCHHLQEEVSRKLLA-DAGEPSRVEMAAS-----IVEV-K-----TSDD-----A
EtnD_ACP-KS_M   80  RLAHLSELDVRAAG-RPVAETERERPALG-----GDIAA-----R
Bat2_ACP-KS_M   74  MLASHLLSQEAFRSRDA-SMNPAPGPTS-GWMATATASDKRNEPVS-----IAFV-KTLAEPQSWAP-----K
Bat3_ACP-KS_M   66  VLAHLLTRGSLRSREA-SEPAQALVATPLATLIQVVECNERVISTA-----ITTL-ETLALPLPLAP-----K
OocN_ACP-KS    57  ALARAVARMQTDAPAKEGRQNA-----
PedF_ACP-KS_2   73  RLAHHIVVEYCHTLDVPAALPCPE-----LSV-SEPAMDIP-----L
PsyD_ACP-KS_2   73  NLTQHQQRIEETRLRQMSGDA-----
OnnI_ACP-KS_2   73  ALADHISKTFQGDINIREKVKETGDLK-----KGITDEPKWEPD-----P
Bond_ACP-KS_M   46  ALAKLLASTGERGHAGAPTAEA---PPA-----PT-----RPADPEP-----SRSPQ
EtnE_ACP-KS     66  BLAYFVEHHAATLSLLCGATAPAAPATAVSA-----G-A-----GPA-----RRRAR
EtnF_ACP-KS     73  GLACHLAEEHREACQRWAGCPDGDAPDD-VARA-E--P-RG---PS-----SAAQ
EtnD_ACP-KS     55  ALAGHLSRHHAAACRWTCMDRLADAPVHANEP-----GGAM-----APLDAGA-----RRDEA
PedI_ACP-KS     59  ALSYLLREHQRITALTCLRPSAGLDQSGRS-----GER-----MV-----SS--
PsyD_ACP-KS     61  DLACHLANTGVEV-----R-----G-----P-----SPIV
OnnB_ACP-KS     60  ALVTHFLQHOREALHALMCIDEFT---DSE-----P-----SEEV
PedF_ACP-KS     60  ALAYLVLNKTVSCRAWTKLRDESVLVADAAR-----RGLP-----LP-----ET--
OnnI_ACP-KS     74  ELVSLLDPCPHACTRWVCRFEAPPVREPEPET-E--TVNGLC---EP-----ASQV
VirA_KS         1  -----
BaeM_ACP-KS     60  BLAYLKKTVSFECE-----RAAEPS-----RSEA
Bat3_ACP-KS     61  KLAAYLVERYPSELALGLKIDDTQRAKQPAV-----VS-----RTTS
OocL_ACP-KS_M   61  ELTEIITITRFQQAARYDPSHRSDGRQPPASAT--KPENGP-----AAGEKSA-----RSIQA
PsyD_ACP-KS_3   75  ALAHLVEYHQTQQMCAGEPEVCD-EFKVE-----A-----
Bat3_ACP-KS_2   61  DFAYYLVEKHGEAFVCFE---PPVQTSGSKE---QP-----VPRAEVP-----VSV--
Ta-1_ACP-KS_2   62  ALAGYLCESHGPRLLDQLCAEAAAFEPASPPS-----AP-----MPVEE
TaO_ACP-KS      62  ALVGRLVSEHRPLAERHRGVGGPVADTSP-E-----DT-----LPTPEVS
BaeJ_ACP-KS     72  GLAGYLAKHE---PGRFGEKKESPKK---E-----QP-----KAQKEKM
BaeN_ACP-KS     72  EFAYHLLISEYEEEFAGRAVAVNKTQVSARPVK-----ET-----QITREAV

```

Ta-1_ACP-KS 78 PAR-----E-----DTSERDYEGSLAVIGLSCQLPGAADPWFVKNLREGRDQVYVRHRELRE--LGV-----PE
 BonA_ACP-KS_2 109 PAP-----ADAPRAR--DAVVRDDEPIAIVGMSGRFPGARNVDFEWTILRDGLDMWRPAPAQGP-----AD
 Bat2_ACP-KS 110 PAD-----SQASTVA--RRGLAPOEFMAIIGMSGRFPGARDVDFEWHVLEGRDAVQEIPLTRDWRNYYCAPDTSVGAQ
 VirA_ACP5b
 OocJ_ACP-KS 89 PPP-----SPRREQAAEDKREHDGDIAVIGMSGRFPGMANVDFEWMVLAAGDCITSEVPAADRWDAAYYD-----ED
 Ta-1_ACP-KS_M 104 VQA-----VQAPAIVKKEVAPAGTVRIAVIGMSGRFPGAEVLRQVWDLNRDGVSCISELPHYTRWDVATA-----K
 EtnE_ACP-KS_M 151 VAC-----VAVRERVIAAGPGEAACRIAVIGMSGRFPGADLDEWVSNLARGVDAVAEVPASRWSVERYYD-----PA
 BaeL_ACP-KS_M 109 -OK-----EPEFVQAKREPKEDEIAIAIAIVGMSGRFPGAPLPELKTWDLNARAKNAIRDIPLSRWDMNKYYD-----PA
 OocL_ACP-KS 82 PAA-----SDAPAEAT--AASHRRDEKIALIGMAGRFPDANLDQVWDLNLANGRNSVREWRFRWRVLDQYED-----AD
 BonA_ACP-KS_M 113 ARP-----AASPVQGA--TPPHREHPIAIVIGMAGRFPGANGMDQVWDLNLAGRDSVEEIPPSRWVADYYD-----AD
 BonA_ACP-KS 106 BIA-----P-----VPPSSIEPIAIVGMSGRFPGAGNRELVAAALAGGEDLIE--PVTWRWDSRHYYA-----D
 PedI_ACP-KS_M 101 FCA-----PP-----VRQPTDIPAIIVIGMSGRFPDAPDKAFWENVEGRDCVSEVPAERWSEEEYYD-----SD
 OnnB_ACP-KS_M 84 PAS-----QPLTPVSKRWLPEQKTGRIAVIGMAGRFPASTLDFWENVAEGRNCVSEVPESRWVDEYYD-----AD
 PsyA_ACP-KS_M 104 EIS-----DVEGDMPAIAIVIGMSGRFPANNVEALWDLNVEGRDCVSEVPLDRWVDAVYD-----PT
 EtnD_ACP-KS_M 116 PAR-----EPEARAP--RPAPEHPIAIVIGMAGRFPKARDLDEFWENLARGVDCVTEIPDRWVAEHHYD-----PR
 Bat2_ACP-KS_M 135 PAS-----PPEELVLS--GTQVLLIGIAGMAGRFPARNAQVFWENLAQGLDCVSEVPSRWSEEEHYD-----PN
 Bat3_ACP-KS_M 128 PAS-----PPEESVP--GTQVLLIGIAGMAGRFPARNAQVFWENLAQGLDCVSEVPSRWSEEEHYD-----PN
 OocN_ACP-KS 80 -----APG-----PLPGSGDIAIVIGMDINVAADNAAEFWQLREGRSSVGOVPOGRQREDLS-----M
 PedF_ACP-KS_2 109 PAV-----QAVPSSLPRREAVVQTIGIAGMAGRFPAGADSDAIVQNVVAGVNPVTESELYLPYHAYS-----P
 PsyD_ACP-KS_2 96 --I-----AEVD-----MSNPSEHPIAIVIGMAGRFPAGADNDAFWENLVQGVDAVGEIPARYLSREAYS-----S
 OnnI_ACP-KS_2 113 PA-----DMSVEFGEKDFGSSPRIAIVIGMSGRFPAGAKMSEFWQNLISGDDTEGELPHHYLPEESHS-----P
 Bond_ACP-KS_M 88 EV-----AASAAPSSLGAPADFEPIAIVIGMSGRFPDANLDEWVWDLNLAGRDCITEIPHRWSDGGHYS-----ED
 EtnE_ACP-KS 109 RLPS-----RGRAGEAREAAIAIVIGMAGRFPDANLDAEFWENLRSKDCITEIPDRWRDHRAYED-----AE
 EtnF_ACP-KS 115 EARS-----GERAPAEALPAQELDAIVIGMAGRFPARTLDEFWENLRAGKDCITEIPDRWPEEGHYD-----GA
 EtnD_ACP-KS 104 ELPK-----SEEAAGARPARPEPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWSDGGHYS-----ED
 PedI_ACP-KS 99 -----SISAPSSIEHPIAIVIGMSGRFPASADLDEWQVLAAGESCISEVPAERWVDEDFH-----ED
 PsyD_ACP-KS 81 TAPR-----LA-VASQLPSRAEVEPIAIVIGMAGRFPDANLDAEFWENLRAGKDCITEIPDRWSDGGHYS-----AS
 OnnB_ACP-KS 92 VAPA-----AS-VRGAPSNRLREPIAIVIGMSGRFPDANLDAEFWENLRAGKDCITEIPDRWPEEGHYL-----AD
 PedF_ACP-KS 100 -APV-----VERNVLVPG--NAVQEPPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWRUENEFH-----ED
 OnnI_ACP-KS 120 BOPH-----ARSAKKSCNSVQEPPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWSDGGHYL-----AD
 VirA_KS 1 -----IAVIGVGRFPADLDRWNLREGRDQVYVRHRELRE--LGV-----PE
 BaeM_ACP-KS 86 BAPD-----TPEHHKNTGDIIVIGMAGRFPRAKSVAEFWENLRAGKDCITEIPKSRWDMKTYK-----NV
 Bat3_ACP-KS 99 SDPE---AGVNAQRTLLSTGEPSOPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPASRWDRWRD-----SL
 OocL_ACP-KS_M 114 LLAKSHQIRQKRQPVVVPANGRYEPIAIVIGMSGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----NP
 PsyD_ACP-KS_3 107 -----LDAQTPVEPVRTGPEPIAIVIGMSGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----DP
 Bat3_ACP-KS_2 101 DAVHRRSRLNLDLPHVPLRQVPEPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----DP
 Ta-1_ACP-KS_2 101 --ARSE---E----RVAPQPRAREPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----DP
 TaO_ACP-KS 102 --RRSGRRSPGFSPAAVKVPQGRPEPIAIVIGMSGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----DP
 BaeJ_ACP-KS 106 --QRKKRFATVMNASAAATQEPRRFPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----DP
 BaeN_ACP-KS 113 --KRRRTLPETL--PQTVQRQGPPEPIAIVIGMAGRFPDANLDEFWENLRAGKDCITEIPDRWDRWRYG-----DP

PKS System	Subunit name	Length of linker following β -modification module	Length of linker following non-modifying module
Myxovirescin	Ta-1	51	24
	Ta-1		46
	TaO		54
Bonkrekic Acid	BonA	78	63
	BonA		46
	BonD		52
Kalimanticin	Bat2	70	58
	Bat3	71	56
	Bat3		51
Oocydin	OocJ		49
	OocL	69	45
	OocN		24
Etnangien	EtnE	91	48
	EtnD	44	56
	EtnF		52
Bacillaene	BaeL	56	
	BaeJ		48
	BaeN		53
	BaeM		30
Pederin	PedI	67	41
	PedF		47
	PedF		48
Onnamide	OnnB	53	41
	OnnI		49
			56
Psymberin	PsyA	41	
	PsyD		26
	PsyD		29
	PsyD		38
		Average: 62.8	Average: 45.6

b

```
DEBS1      1  - SHFAEENKAL - LVRNHFAVLGHAARVADQAFELGDSLSALELRNRLATGVRLE - T VFD P VV LA HL AEL GATGAEQAAPA
DEBS3      1  LAALSTARRREH AHTREAVAVLGHDDAAIDRRRAFRLGFDSTAVILRNRLAATGCEAA - TLVFDPTTTRTADHYLERVGAEEAQAPAL
DEBS2      1  LAGLSPDQQENLELVANAIAEVLGHESDAEINVRRARSFLGDSLNAIRRRRLSASTGLRLPA - TLVFDPTTTLAQLRRARLGDADQAVVVV
AmphB      1  - QGVSAAEQDRILSLVRETAAVLGHHSPTETERRRAFRLGFDSTAVILRNRLVAATGLRLPS - TLVFDPTPNAVLAFFKATAGTGTAGDPT
PIKS1      1  -----TEITLGLVRAQAQAVLRMSPEWAAADRAFQDGFDSLAVELRNRLTRATGLRLPA - TLVFDPTPLALVSLRSEFLGDEETADARRS
RAPA       1  -----DLVRSSTIDLGLSADRYAPDTSREMCI DSLTAVELRNSLAKATGLRLPA - TLVFDPTPAVLVRLGELFTG - BSPAPERRA -
RAPB       1  -----PEERAKALLKIVRDTAAVVLGHDDARTPVTGAFRLGIDSLTAVELRNSLAKATGLRLPA - TLVFDPTPAVLAARLGEELFTG - ENPVLVET -
RAPC       1  -----ALLKLVSDGAAVVLGHDDTSTIPATTAFKVLGINSLTAVELRNSLAKATGLRLPA - TLVFDPTPALAARLGEELFTG - ENPVVVEGP
PIKS2      1  LGALTGAQQRRVQELVREHVAVVLGHHSPEAVDTGAFRLGFDLSLTAVELRNLKATGLRLPA - TLVFDPTPRLVAFGLAEIIGEQAGVGEQLP
AmphC      1  -----RREALLCLVRTQAACVVLGHDDACQEAARQFQDLGFDLSLTAVELRNLNSATGLRLPA - TLVFDPTPDLADHLSDELIGTGDNIPLPTP
AmphI      1  LARLDAESRDELIDLVRSQVAVLGHHTAEIDAGRAFRLGFDLSLTAVELRNLTVGLRLPA - TLVFDPTPSLDAVYLDDELIGTEVEVAVV -
MtaB       1  VGASPRRLRARLLSLVQVAAVVLGLPPSAPDPRRLQELGIDSLMAIFRNSLAASLRSLLPA - TLVFDPTTLEEFFEARLTPASEASPAISR
AjuB       1  LEAAPAAERMEVIAFVVGQAARVGLGFDGSHPLERKMFELGFDLSLAVELRNROARVGDALPLSNITVFDPTTTHALAGVAGRVKGE - AVI - -
AjuG       1  -----LATHVASEVARVLGFDFAKRPDPAALGDFELGMDSLSLVELRRLQKSLCALPT - TLVFDPTVSDALTAHARNVLDIGRAPPP - -
consensus  1  . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . *
```

```
DEBS1      98  -----TTVPVDEPTIAIVGMACRLPG-V-SPELW-LITSG-DSPAENF-DRGWVLDLVA-D-LGT-TLRAAGFMIG-AG-FDAAEFF
DEBS3      99  V-----REVPKDAEPTIAIVGMACRLPGGVHMCGLLWESVGGGDAIEMFDRGWVDLIDDDPDRRGTVYSRSGAFIDG - AAQFDAEFF
DEBS2      99  -----GADSESPPIAIVGICRFPGGIGSPQLWRLVLAGCANLITGFPADRGWDLGRLHPPDPPNRGRTVYDKGGFLD - AAQFDAEFF
AmphB      98  A-----AVTAADDTIVVGMSCRLPGGANPEETIRLALGADVTSRFPADRGWDAHLGYDDPPDRCGRTVYSVGGFLH - AAQFDAEFF
PIKS1      90  AALPATVGAGAGAGAGTDAADDPIAIVAMSCRYPGDIHSPEDLWRLVMSGGGGITRFPDRGWDLGLYDDPPDAGCRVYSVGGFLH - AAQFDAEFF
RAPA       83  V-----S--VVGQDEPTIAIVGMACRLPGGVSSPEDLWRLVMSGGDAISGFFDRGWVDLGLDDPPDASCKVYCVGGFLDT - AAQFDAEFF
RAPB       92  A-----S--VVGQDEPTIAIVGMACRLPGGVSSPEDLWRLVMSGGDAISGFFDRGWDAHSLDDPPDPAVKVYCVGGFLDS - AAQFDAEFF
RAPC       87  V-----S--VAQDEPTIAIVGMACRLPGGVSSPEDLWRLVMSGGDAISGFFDRGWVGNLYDM - - - - - ACKHRAEGGFLDA - AAQFDAEFF
PIKS2      99  V-----D--GVVDEPTIAIVGMACRLPGGVSSPEDLWRLVMSGGDAISGFFDRGWVGLYDPPDPSRNYVYRAGGFLD - AAQFDAEFF
AmphC      91  A-----GALPPTDDPIAIVGMACRLPGGVSSPEDLWRLVMSGGDATGFPFVNRGWVGNLYDPPDRGHVYRAGGFLH - AAGFDAEFF
AmphI      97  Q-----RGTAVADPIAIVGMACRLPGGVSSPEDLWRLVMSGGDAISGFFVNRGWVGNLYHPPDHPGTAVYSRGGFLD - AAGFDAEFF
MtaB       98  RP-----VKLPARDREGAEPIAIVGMACRLPGSVRSPEELWRLVMSGGDAISGFFVNRGWVGNLYHPPDHPGTAVYSRGGFLD - VDFDASFF
AjuB       96  -----ASRIQASDDPIAIVGICRFPGGADGFEAFVQVLRVGLDVAIVEVREKRWDTARFHDPPDPAKGMYSRGGFLR - IDQFDAEFF
AjuG       85  -----ITAAITRRDEPPIAIVGMACRLPGGANVLAWRLLRGCADATHEVERDR - WLDVAVYDPPDPAKGMYSRGGFVOGIVGDVDFDASFF
consensus  101 . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . *
```

Fig. S9 Analysis of ACP-KS linkers. (a) Portion of a multiple alignment of the ACP-KS regions (junction between two modules) from a selection of *trans*-AT PKS subunits, to determine the lengths (shown in the table) of the linkers intervening between the conserved ACP and KS domains. The C-terminal and N-terminal boundaries of the ACP and KS domains, respectively, were determined by alignment with VirA module 5 KS and ACP_{5b}. Linkers which follow modules where β -modification occurs are indicated by M (2 indicates a second junction from the same subunit). Although only 11 sequences of such linkers are available in the database (that is, both the gene sequence and the product structure are known, validating the β -modification), these linkers are generally longer than comparable regions from the same systems, and the average length of all these linkers (ca. 63 residues) is significantly greater than the average length of linkers separating non-modifying modules (ca. 46). Key: Ta, myxovirescin; Bon, bonkrekic acid; Bat, kalimantcin; Vir, virginiamycin; Ooc, oocydin; Etn, etnangien; Bae, bacillaene from *Bacillus amyloliquefaciens*; Ped, pederin; Onn, onnamide; Psy, psymblerin. (b) Multiple alignment of the ACP-KS regions from a selection of *cis*-AT PKS from both *Streptomyces* and myxobacteria. The C-terminal domain boundary of the ACPs was based on the NMR solution structure of DEBS ACP₂.⁷ The average linker length is 20 residues. Key: DEBS, erythromycin; Amph, amphotericin; PIKS, pikromycin; RAP, rapamycin; Mta, myxothiazol; Aju, ajudazol.

Supplementary Tables

Table S1 Summary of SAXS Data

Construct	R_g (Guinier) (Å) ^a	R_g (GNOM) (Å) ^a	D_{max} (GNOM) (Å) ^b	χ^2 (DAMMIN) ^c	χ^2 (SASREF) ^c	χ^2 (MONSA) ^c
KS	33.1 ± 0.5	32.6	105	1.689	1.333	1.421
ACP _{5a} -ACP _{5b}	29.0 ± 0.2	30.7	103	1.785	24.981	nd
KS-ACP _{5a} -ACP _{5b}	56.5 ± 0.7	60.9	230	2.813	nd	1.885
KS-ACP _{5a} -ACP _{5b} -DD	69 ± 1	72.5	279	2.204	nd	nd
ACP _{5b} -DD	37.0 ± 0.3	37.7	140	1.231	nd	nd
ACP _{5a} -ACP _{5b} -DD	43.5 ± 0.2	46.5	170	1.366	nd	nd

^a R_g is the radius of gyration derived from the atomic models using the program PRIMUS. For experimental data, R_g is given by the Guinier approximation⁸

^b D_{max} is the maximal particle diameter derived from the distance distribution function ($P(r)$) using the program GNOM⁹

^c χ^2 is the discrepancy between the experimental SAXS profile and the theoretical calculated curve using the programs DAMMIN,⁹ SASREF¹⁰ or MONSA¹⁰

nd: not determined

Table S2 NMR and Refinement Statistics for ACP_{5a}

	Protein
NMR distance and dihedral constraints	ACP _{5a}
Distance constraints	
Total NOE	1968
Intra-residue	374
Inter-residue	1594
Sequential ($ i - j = 1$)	490
Medium-range ($ i - j < 4$)	566
Long-range ($ i - j > 5$)	538
Intermolecular	n/a
Hydrogen bonds	0
Total dihedral angle restraints	114
ϕ	57
ψ	57
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.092 ± 0.012
Dihedral angle constraints (°)	0 ± 0
Max. dihedral angle violation (°)	0
Max. distance constraint violation (Å)	0.13
Deviations from idealized geometry	
Bond lengths (Å)	0.010 ± 0.001
Bond angles (°)	2.062 ± 0.018
Improper (°)	n/d
Average pairwise r.m.s. deviation** (Å)	6734–6811
Heavy	0.94 ± 0.10
Backbone	0.27 ± 0.06

**Pairwise r.m.s. deviation was calculated using 20 refined structures. n/a = not available; n/d = not determined.

Table S3 NMR and Refinement Statistics for ACP_{5b}

	Protein
NMR distance and dihedral constraints	ACP _{5b}
Distance constraints	
Total NOE	1846
Intra-residue	371
Inter-residue	1475
Sequential ($ i - j = 1$)	500
Medium-range ($ i - j < 4$)	522
Long-range ($ i - j > 5$)	453
Intermolecular	n/a
Hydrogen bonds	0
Total dihedral angle restraints	110
ϕ	55
ψ	55
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.096 ± 0.023
Dihedral angle constraints (°)	0 ± 0
Max. dihedral angle violation (°)	0
Max. distance constraint violation (Å)	0.17
Deviations from idealized geometry	
Bond lengths (Å)	0.010 ± 0.001
Bond angles (°)	2.054 ± 0.018
Improper (°)	n/d
Average pairwise r.m.s. deviation** (Å)	6836–6909
Heavy	1.15 ± 0.15
Backbone	0.38 ± 0.07

**Pairwise r.m.s. deviation was calculated among 20 refined structures. n/a = not available; n/d = not determined.

Supplementary References

- 1 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W.H. Gerwick, D. H. Sherman, D.H., and Smith, J.L., *Chem. Biol.*, 2013, **20**, 1340–1351.
- 2 Y. Tang, C. -Y. Kim, I. I. Mathews, D. E. Cane, and C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 11124–11129.
- 3 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher, and C. Hertweck, *Nature*, 2013, **502**, 124–128.
- 4 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, and V. N. Uversky, *J. Mol. Biol.*, 2006, **362**, 1043–1059.
- 5 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky, *Biochem. Biophys. Acta*, 2010, **1804**, 996–1010.
- 6 H. P. Erickson, *Biol. Proced. Online*, 2009, **11**, 32–51.
- 7 V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi, and C. Khosla, *Protein Sci.*, 2007, **16**, 2093–2107.
- 8 P. V. Konarev, V. V. Volkov, A.V. Sokolova, M. H. J. Koch, and D. I. Svergun, *J. Appl. Crystallog.*, 2003, **36**, 1277–1282.
- 9 D. I. Svergun, *J. Appl. Crystallog.*, 1992, **25**, 495–503.
- 10 M. V. Petoukhov, and D. I. Svergun, *Biophys. J.*, 2005, **89**, 1237–1250.

II. Caractérisation d'une nouvelle classe d'interface de docking chez une PKS de type *trans*-AT

Dorival, J., Annaval, T., Risser, F., Collin, S., Roblin, P., Jacob, C., Gruez, A., Chagot, B. et Weissman, K. J. Characterization of Intersubunit Communication in the Virginiamycin *trans*-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **138**, 4155–4167 (2016).

Après l'étude du module 5, nous nous sommes intéressés de plus près au rôle que pouvait jouer le domaine en C-terminale du module 5 en se basant sur l'hypothèse qu'il pouvait s'agir d'un domaine de docking (Broadhurst, *et al.*, 2003 ; Buchholz, *et al.*, 2009 ; Whicher, *et al.*, 2013). Nous avons alors identifié un domaine analogue en N-terminale de la sous-unité VirFG qui porte le module 6 et le module 7 de la PKS. Nous avons tout d'abord produit chacun de ces DD potentiels pour caractériser leur interaction, tout d'abord par gel filtration analytique (S75 10/300) puis par RMN (HSQC ^1H - ^{15}N). L'interaction a ensuite été mesurée par ITC, ce qui a permis de déterminer un K_d de $5,8 \pm 0,2 \mu\text{M}$. Une affinité similaire a également été mesurée entre l'ACP-DD VirA et le DD VirFG ainsi qu'entre le DD VirA et le DD-*linker* VirFG (DD VirFG avec le *linker* entre le DD VirFG et le KS₆ – la construction incorporant le KS₆ étant non soluble. Ceci montre que les domaines/*linkers* flanquant ne semblent pas intervenir dans l'interaction entre les sous unités.

Nous avons ensuite résolu la structure RMN d'une protéine de fusion entre les deux DD liés par un *linker* de 8 résidus (Whicher, *et al.*). La structure révèle que les DD sont composés chacun de deux hélices qui interagissent entre elles avec une topologie unique pour une interaction protéine-protéine, ce qui dénote une différence considérable entre les interfaces de docking chez les PKS de type *cis* et chez les PKS de type *trans*. Par ailleurs, de nombreux points laissent penser que le DD VirFG seul est une protéine intrinsèquement désordonnée. En effet, les spectres de dichroïsme circulaire (CD) sont caractéristiques d'une protéine désordonnée pour le DD VirFG alors que le DD VirA est structuré en hélice α . Une mesure en double cuve a permis de mettre en évidence un repliement induit du DD VirFG en présence de son partenaire. Cette analyse a été confortée par des expériences en SAXS qui montre que le DD VirFG n'est pas une protéine globulaire : le Kratky plot est caractéristique

d'une protéine plutôt allongée, donc probablement désordonnée, présentant cependant quelques éléments de structures secondaires.

Grâce à la structure RMN, nous avons mis en évidence trois couples de résidus polaires importants pour l'interaction. La nature polaire de ces résidus est de plus conservées parmi les DD pouvant appartenir à la même classe que VirA/VirFG, bien que les charges peuvent varier. Nous avons ensuite muté chacun des 6 résidus polaires identifiés. Nous avons constaté des variations dans les spectres de dichroïsme circulaire pour les mutants du DD VirFG. Cependant, il était impossible de savoir si les mutations du DD VirFG entraînaient une perturbation de l'interaction, ou bien de la capacité à se replier en présence du partenaire. Ces trois mutants n'ont donc pas été utilisés. Après vérification de l'intégrité de la structure des trois mutants du DD VirA par CD, nous avons noté une diminution de l'affinité par chacune de ces mutations, mesuré avec le partenaire DD VirFG type sauvage. Nous avons donc pu conclure que l'interaction était assurée par ces interactions polaires, mais elle est aussi vraisemblablement stabilisée par un cluster d'acides aminés hydrophobes (L6939, L6943, V6946, I6953, A6956, V6957, et L6960 pour le DD VirA et A3, L7, F10, A20, et L24 pour le DD VirFG). La surface d'interaction entre les deux DD est d'environ 730 Å², ce qui est comparable avec les autres classes de DD (environ 1000 à 1100 Å² pour les DD de type 1a et 1b respectivement, et environ 1800 Å² pour les DD de type 2). En effet, l'interface entre les DD de VirA/VirFG est monomérique, et l'ensemble de la PKS est dimérique. Deux interfaces de docking se formeraient donc simultanément (la surface d'interaction serait alors de 1460 Å environ). Pour les autres classes de DD, une seule interface implique deux copies de chaque DD. L'affinité entre les DD de chacune des classes est également comparable et assez faible dans tous les cas (K_d entre 2 et 50 μM).

Des expériences par SAXS ont également été menées sur différentes constructions afin d'identifier la position des domaines flanquants. Ainsi, nous avons résolu la structure à faible résolution de la construction ACP_{5b}-DD_A-DD_{FG}-KS₆-*linker*. Les domaines de docking adoptent une position centrale et sont entourés de part et d'autre par l'ACP_{5b} et le KS₆. Nous avons également pu proposer un modèle d'interaction entre le module 5 et le module 6. Ce dernier se base sur les données SAXS obtenus précédemment sur le tridomaine KS₅-ACP_{5a}-ACP_{5b} et celles de la

fusion ACP_{5b}-DD_A-DD_{FG}-KS₆-*linker*. En effet, les deux enveloppes se superposent parfaitement au niveau de l'ACP_{5b}, commun aux deux constructions.

Dans ce travail collaboratif au sein de l'équipe, j'ai participé à la résolution de la structure par RMN (attribution et structure du DD de VirA marqué en interaction avec le DD VirFG non marqué ; le Dr. Thibault Annaval et le Dr. B. Chagot ont résolu la structure de la protéine de fusion entre les deux DD), et j'ai réalisé la partie SAXS avec le Dr. Arnaud Gruez. La purification des DD et des mutants a été partagée entre le Dr. T. Annaval et moi-même, Sabrina Collin nous a aidé pour la purification de la fusion ACP_{5b}-DD_A-DD_{FG}-KS₆-*linker*, et les spectres de CD et les expériences d'ITC ont été réalisées par Fanny Risser.

Characterization of Intersubunit Communication in the Virginiamycin *trans*-Acyl Transferase Polyketide Synthase

Jonathan Dorival,^{†,⊥} Thibault Annaïval,^{†,⊥} Fanny Risser,[†] Sabrina Collin,[†] Pierre Roblin,^{‡,§} Christophe Jacob,[†] Arnaud Gruez,^{*,†} Benjamin Chagot,^{*,†} and Kira J. Weissman^{*,†}

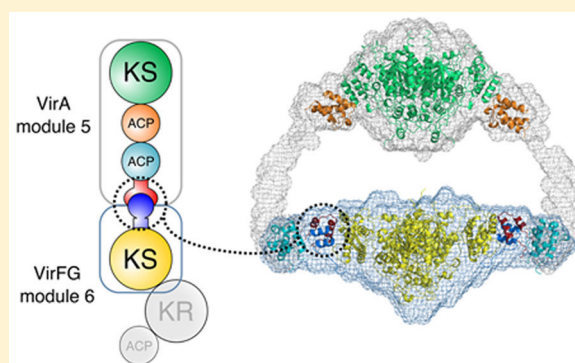
[†]UMR 736S, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandœuvre-lès-Nancy CEDEX, France

[‡]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 Gif-sur-Yvette CEDEX, France

[§]UR1268 Biopolymères, Interactions Assemblages (BIA), INRA, Rue de la Géraudière BP 71627, 44316 Nantes CEDEX 3, France

Supporting Information

ABSTRACT: 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 (Vir) system. In the absence of C-terminal partner (VirA^{CDD}) or a downstream catalytic domain, the N-terminal DD (VirFG^{NDD}) exhibits multiple characteristics of an intrinsically disordered protein. Fusion of the two docking domains results in a stable fold for VirFG^{NDD} and an overall protein–protein complex of unique topology whose structure we support by site-directed mutagenesis. Furthermore, using small-angle X-ray scattering (SAXS), the positions of the flanking acyl carrier protein and ketosynthase domains have been identified, allowing modeling of the complete intersubunit interface.



INTRODUCTION

Synthesis of structurally complex, reduced polyketides in bacteria occurs on gigantic multienzymes called polyketide synthases (PKSs).¹ These systems are likened to molecular-scale assembly lines because 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.² *trans*-AT PKSs, however, are characterized by one or more free-standing enzyme activities (including notably the ATs) that collaborate in *trans* with the PKS multienzymes to accomplish both chain building and processing (Figure 1).^{3,4} They also typically include modules of nonribosomal peptide synthetase (NRPS) that incorporate amino acids into the growing intermediates.

One conserved feature of *cis*- and *trans*-AT PKSs is the distribution of the functional modules among multiple polypeptide subunits.⁵ 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 because no modifications are made to the component modules. Indeed, modules are increasingly viewed as discrete operational units, given the high interdependence of the functional domains.^{6,7}

In *cis*-AT PKSs, 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.^{8–12} Three evolutionarily related docking domain classes have been identified to date in such PKS systems,¹³ as well as a fourth class which operates at NRPS–NRPS/–PKS junctions in mixed PKS–NRPSs.¹⁴ The NMR or X-ray crystal structures of covalently fused complexes of the PKS docking domains have been solved^{15–17} (Figure 2), and the N-terminal DDs were shown to adopt the same structures when expressed together with the downstream ketosynthase (KS) domains.^{17–19} The close structural homology between the two actinobacterial classes of DDs prompted their grouping into Class 1 (referred to here as 1a and 1b), whereas the third class, which is present

Received: December 22, 2015

Published: March 16, 2016

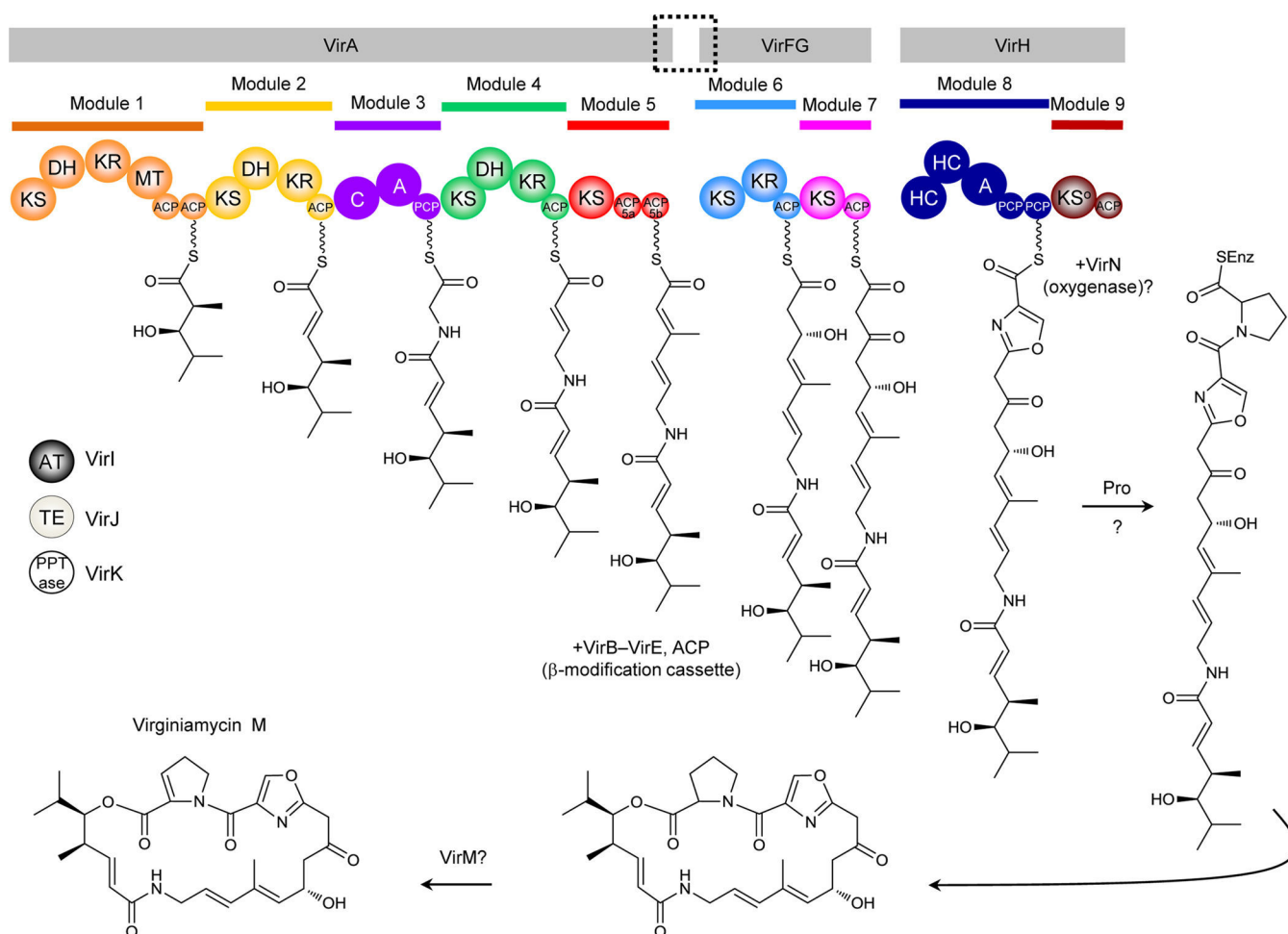


Figure 1. Virginiamycin (Vir) *trans*-AT polyketide synthase. The PKS comprises three subunits, VirA, VirFG (whose N-terminal sequence was recently revised by us),²¹ and VirH as well as additional modules that were not identified in the initial cluster description⁴ for starter unit selection and introduction 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), nonribosomal peptide synthetase modules (3 and 8), an inactive domain (KS^o of module 9), and a cassette of enzymes that together introduce a β -methyl functionality. The experiments described here reveal the molecular details of the interface between subunits VirA and VirFG (boxed).

in myxo- and cyanobacteria, has been designated as Class 2.¹⁷ (By extension, the mixed PKS-NRPS DDs constitute Class 3.)

Class 1 and 2 DDs share two central features: the N-terminal docking domains are homodimeric and form a coiled-coil motif that serves as a platform for docking of the likewise α -helical C-terminal docking domains. The docking interfaces are largely hydrophobic in character and complementary in shape, whereas 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 (Figure 2A,B) incorporate two α helices upstream of the docking α helix which together form a dimerization motif. Class 2 C-terminal docking domains, however, comprise only two α helices, both of which associate with the N-terminal docking domain. Furthermore, the coiled-coil motif of the Class 2 N-terminal DD is preceded by an additional α helix; thus, docking gives rise to an overall eight α -helical bundle (Figure 2C), in contrast to the four α -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 μ M range.^{16,17,20}

We recently characterized by small-angle X-ray scattering (SAXS)²¹ the structure of *apo* module 5 from the VirA subunit of the virginiamycin hybrid *trans*-AT PKS-NRPS⁴ of *Streptomyces virginiae* (Figure 1). In the course of this work, we noted that the extreme C-terminal sequence (designated “VirA^{CDD}”) showed similarity to a region from the CurA subunit of the curacin *cis*-AT PKS, which had been described as a dimerization motif.²² 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 because the module retained its open form. This prompted us to investigate whether VirA^{CDD} was rather a docking domain, mediating a specific interaction with a partner DD located at the N-terminus of subunit VirFG (“VirFG^{NDD}”). We show here that VirA^{CDD} and VirFG^{NDD} 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

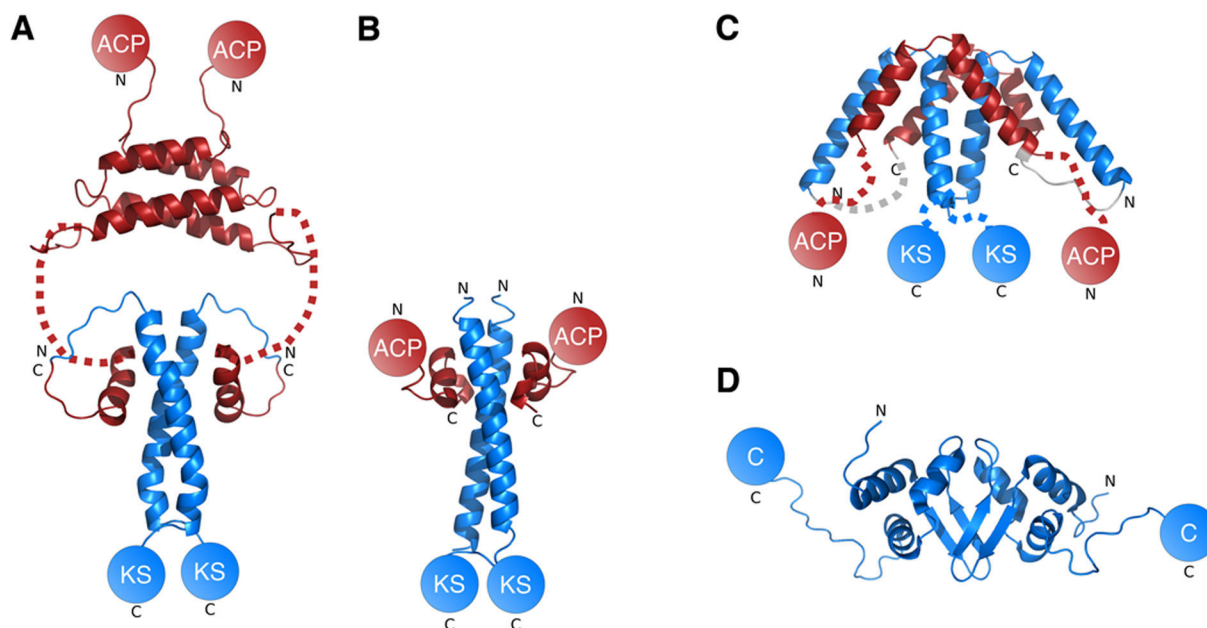


Figure 2. Three classes of docking domains from *cis*-AT PKSs. (A) NMR structure of a complex of covalently fused Class 1a docking domains (Protein Databank (PDB): 1PZQ, 1PZR).¹⁵ The C-terminal DDs (red) comprise three α helices, of which the first two form a four α -helical bundle that acts as a dimerization element. The third helix wraps around the coiled-coil motif formed by the partner N-terminal docking domain (blue). The long, flexible linker joining the second and third α 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).¹⁶ 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 α 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).¹⁷ The C-terminal DD does not incorporate a dimerization region. Instead, its two α helices associate with two α helices provided by the N-terminal docking domain to form an overall 8 α -helical bundle. (D) NMR structure of a class 3 N-terminal docking domain (PDB: 2JUG).¹⁴ Each protomer within the DD homodimer comprises three α helices and two β strands, which adopt an overall $\alpha\beta\alpha\alpha$ 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; and C, condensation.

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 compared to that of *cis*-AT PKSs.

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/>) with secondary structure prediction by PsiPred²³ and displayed using BoxShade (http://www.ch.embnet.org/software/BOX_form.html), or using PROMALS3D.²⁴ The C-terminal boundaries of the ACP domains were determined by multiple sequence alignment of the subunit ends and comparison with domain boundaries identified in the solved structures of ACPs 5a and 5b from the virginiamycin PKS.²¹ 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,^{18,19} condensation (C),²⁵ and enoyl reductase (ER)²⁶ domains.

Materials and DNA Manipulation and Sequencing. *S. virginiae* genomic DNA was obtained from the National Institute of Agrobiological Sciences Genebank, Japan. 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 *Escherichia coli* was carried out according to standard methods.²⁷ 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^CDD, VirFG^NDD, VirFG^NDD-linker, VirA ACP_{5b}-^CDD, and VirFG^NDD-KS₆) were amplified from *S. virginiae* genomic DNA using forward and reverse primers incorporating BamHI and HindIII restriction sites, respectively (Table S1), and were ligated into the equivalent sites of vector pBG-102 (Center for Structural Biology, Vanderbilt University); in the case of VirFG^NDD-KS₆, an internal BamHI 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 His₆-SUMO tag. Following cleavage of the SUMO-His₆ 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^CDD–VirFG^NDD, VirA ACP_{5b}-^CDD–VirFG^NDD, and VirA ACP_{5b}-^CDD–VirFG^NDD-KS₆) were constructed using overlap PCR. Specifically, fragments encoding VirA^CDD, VirFG^NDD, VirA ACP_{5b}-^CDD, and VirFG^NDD-KS₆ were amplified and then used as templates in a second PCR to amplify the fusion construct, and the resulting fragments were ligated into the vector pBG-102. Site-directed mutants of VirA^CDD and VirFG^NDD were obtained by site-directed mutagenesis using Phusion high-fidelity DNA polymerase from vector pBG-102 containing VirA^CDD and VirFG^NDD wild-type sequences.

The vectors were used to transform *E. coli* BL21 (DE3), and constructs were expressed as His₆-SUMO-tagged fusions by growth in LB medium at 37 °C to an A₆₀₀ 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^{CDD}, VirA ACP_{sb}-^{CDD}, VirA ACP_{sb}-^{CDD}-VirFG^{NDD} (covalent fusion), and all VirA^{CDD} mutants and in buffer 2 (50 mM Tris-HCl (pH 7.5), 150 mM NaCl) in the case of VirFG^{NDD}, VirFG^{NDD}-linker, VirA^{CDD}-VirFG^{NDD} (covalent fusion), and all VirFG^{NDD} mutants, and in buffer 3 (50 mM sodium phosphate (pH 7.5), 400 mM NaCl, 10% glycerol) for VirA ACP_{sb}-^{CDD}-VirFG^{NDD}-KS₆ (covalent fusion). Cells were lysed by sonication, and cell debris was 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_{sb}-^{CDD}-VirFG^{NDD}-KS₆ 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 His₆-SUMO-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_{sb}-^{CDD}-VirFG^{NDD}-KS₆ (covalent fusion) in 20 mM sodium phosphate (pH 7.5), 250 mM NaCl, 2 mM EDTA, and 5% glycerol resulted in a homogeneous preparation of each protein. Production of ¹⁵N- or ¹³C,¹⁵N-enriched VirA^{CDD}, VirFG^{NDD}, and VirA^{CDD}-VirFG^{NDD} (covalent fusion) for NMR structure elucidation was carried out by growth in M9 minimal medium. The minimal medium was supplemented with ¹⁵NH₄Cl (0.5 g/L) and ¹³C₆-glucose (2.0 g/L) as the sole sources of nitrogen and carbon. Isotopically-labeled chemicals were purchased from Eurisotop.

Analytical Gel Filtration. VirA^{CDD} and VirFG^{NDD} (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 reinjected. Calibration was carried out using Gel Filtration Standard (Bio-Rad).

Size-Exclusion Chromatography Multiangle Laser Light Scattering. The oligomeric state of multiple constructs was determined by size-exclusion chromatography multiangle laser light scattering (SEC-MALLS) by the Macromolecular Interaction Platform of I2BC (UMR 9198, <http://www.i2bc.paris-saclay.fr>). For this, SEC was first carried out on a Superdex 200 10/300 GL Increase column (GE Healthcare Life Sciences) at 17 °C in phosphate buffer (50 mM NaPi (pH 7.5), 250 mM NaCl, 5% glycerol) using a Shimadzu Prominence HPLC. Multiangle laser light scattering (MALLS) was measured using a MiniDAWN TREOS (Wyatt Technology) equipped with a QELS module (Wyatt Technology). Refractometry was monitored using an Optilab T-rEX (Wyatt Technology). Data processing was carried out with ASTRA 6 software (Wyatt Technology) using a refractive-index increment of 0.183 mL/g (dn/dc) to calculate the protein concentration from the refractive-index measurements.

Isothermal Titration Calorimetry. ITC was performed using a MicroCal ITC 200 (GE) at 20 °C. A 300 μL aliquot of VirA^{CDD} (wild type or mutant) at 60 μM was placed in the calorimeter cell, and VirFG^{NDD} (wild type or mutant at 600 μM) or VirFG^{NDD}-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 at 120 s intervals. For the experiment with VirA ACP_{sb}-^{CDD}, the construct (900 μM) was added to VirFG^{NDD} (60 μM), as described above. The heat of reaction per injection (microcalories/second) was determined by

integration of the peak areas using either the Origin (MicroCal) software or Nitpic.²⁸ These software provide the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N), and the dissociation constant (K_d). The heats of dilution were determined by injecting VirA^{CDD} and VirFG^{NDD} 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^{CDD} was undergoing homodimerization as the concentration in the calorimeter cell increased because a change in the heat signal was detected up to a concentration of 30 μM. No such change was observed over the whole concentration range for VirFG^{NDD} (up to 78 μM).

NMR Data Acquisition. To minimize the amount of protein needed, 300 μL of protein solution were loaded into a 4 mm NMR tube. ¹⁵N-¹H HSQC spectra of VirA^{CDD} were recorded at a concentration of 100 μM in the absence of partner and in the presence of 200–400 μM ¹⁴N VirFG^{NDD}. ¹⁵N-¹H HSQC spectra of VirFG^{NDD} were recorded at a concentration of 100 μM either alone or in the presence of 400 μM ¹⁴N VirA^{CDD} or 200 μM ¹⁴N VirA ACP_{sb}-^{CDD}. Heteronuclear spectra were acquired on ¹³C,¹⁵N-VirA^{CDD} at a concentration of 1.2 mM in the presence of a 2-fold excess of ¹²C,¹⁴N-VirFG and on ¹³C,¹⁵N-VirA^{CDD}-VirFG^{NDD} at a concentration of 0.9 mM. All NMR data were recorded at 17 °C on a Bruker DRX600 spectrometer equipped with a cryogenic probe (Plateforme de Biophysicochimie Biologie Structurale–FR 3209 Université de Lorraine/CNRS).

Backbone and sequential resonance assignments were obtained by the combined use of 2D ¹⁵N-¹H and ¹³C-¹H HSQC spectra and 3D HNCA, HNCACB, CBCA(CO)NH, HNHA, HN(CA)CO, and HNCO experiments. Assignments of aliphatic side chain resonance were based on 2D aromatic ¹³C-¹H HSQC, 3D (H)CC(CO)NH, H(CC) (CO)NH, HCCH-COSY, and HCCH-TOCSY experiments (reviewed in ref 29). To collect NOE-based distance restraints for the VirA^{CDD}-VirFG^{NDD} structure calculations, 3D ¹⁵N NOESY-HSQC and ¹³C NOESY-HSQC were recorded on uniformly ¹³C,¹⁵N-enriched VirA^{CDD}-VirFG^{NDD} 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.³⁰ 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.³¹ In each cycle, 100 structures starting from random torsion angle values were calculated with 10 000 steps of torsion angle dynamics-driven simulated annealing. A total of 478 NOE-based distance and 54 backbone angle restraints were used for structure calculation of the VirA^{CDD} domain, whereas 1372 NOE-based distance and 98 backbone angle restraints were used for VirA^{CDD}-VirFG^{NDD}. The angle restraints were obtained from ¹³Ca, ¹³Cβ, ¹³C', ¹⁵N, ¹HN, and ¹Hα chemical shifts using TALOS+³² with an assigned minimum range of ±20°. 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 (refs 33 and 34) 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 mol⁻¹ Å⁻² and 50 kcal mol⁻¹ rad⁻², respectively. The final VirA^{CDD}-VirFG^{NDD} representative

ensembles correspond to the 20 conformers from each calculation with the lowest restraint energy terms.

Analysis by Circular Dichroism. Circular dichroism (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 μL of protein sample at 100 μM was used for all the measurements except that of 10 μM VirA^{CDD} (100 μL 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. 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^{CDD} into one cuvette and VirFG^{NDD} into another cuvette (30 μL each at 100 μM) and analyzing the two cuvettes simultaneously. The same result was obtained by averaging the spectra of VirA^{CDD} and VirFG^{NDD} acquired separately. We then combined the two docking domains, and placed the mixture into both cuvettes (30 μL at 100 μM), again analyzing the two simultaneously. Spectrum deconvolution was carried out using the CDNN software.³⁵

Small-Angle X-ray Scattering Data Collection. 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 to 0.5 \AA^{-1} . The scattering vector is defined as $4\pi/\lambda \sin \theta$, where 2θ 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.³⁶ 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,^{37,38} presumably by stabilizing the structure of the proteins. A 50 μL aliquot of the protein sample (at 10 or 15 mg/mL) was then injected, and complete data sets were acquired. The protein concentration downstream of the elution column was followed via the absorbance at 280 nm with an *in situ* spectrophotometer. In contrast to classical SAXS experiments that are conducted in batch using several protein concentrations within a standard range (e.g., 0.1–10 mg/mL⁻¹), the fact that data collection is coupled to a gel-filtration column means that analysis of the required multiple concentrations of the protein occurs within a single experiment because many different positions within the elution peak are sampled during the course of the measurement (typically 50–100 frames are acquired).³⁶ Data reduction to absolute units, frame averaging, and subtraction were performed using FOXTROT, a dedicated in-house application. Each acquisition frame of the experiment yielded a scattering spectrum, which was then analyzed using FOXTROT, giving a R_g (radius of gyration) as well as an $I(0)$ value, which depends on the protein concentration at that location in the elution peak from the Guinier law (approximation $I(q) = I(0) \exp(-q^2 R_g^2/3)$ for $qR_g < 1.3$). It is thus the fact that the concentration naturally varies across a gel-filtration peak that provides a full range of dilutions as used in classical SAXS experiments. Under our experimental conditions (buffer, dilution over the gel filtration step, etc.), the fact that we observed a constant R_g for a significant proportion of the concentrations present in our gel filtration peaks means that these measurements were independent of concentration and thus that we were effectively under conditions of infinite dilution. (See Supporting Information Methods for a model data set.) In the following step of data analysis, all the frames that exhibited the same R_g as a function of $I(0)$ were corrected for buffer signal and averaged, thus ensuring that the data reflected only the signal arising from the protein structure and not from intermolecular

interactions. The distance distribution function $P(r)$ and the maximum particle diameter D_{max} were calculated by Fourier inversion of the scattering intensity $I(q)$ using GNOM.³⁹

Determination of Construct Molecular Weights. Classically, molecular weights can be derived from the SAXS data using the $I(0)$ and the known protein concentration. Although in our experimental setup the protein concentration is known when the peak elutes from the gel-filtration column, the fact that there is a delay between exiting the column and the acquisition of the SAXS data means that the concentration could have changed prior to the measurement. Therefore, to determine the molecular weights of our constructs, we relied on two alternative methods: SAXS MoW,⁴⁰ which relies on integration of the distance distribution function (because this is known to be proportional to molecular weight), and size-exclusion chromatography multiangle laser light scattering (SEC-MALLS), for which the intensity of light scattered off protein molecules in solution is proportional to their concentration and molecular weight.⁴¹

Ab Initio Shape Determination. *Ab initio* protein shapes were calculated from the experimental SAXS data using the bead-modeling program DAMMIN.⁴² At least 20 independent fits were carried out without imposition of symmetry for the monomeric VirFG^{NDD}, the VirA^{CDD}/VirFG^{NDD} and VirA^{ACP_{sb}-CDD}/VirFG^{NDD} non-covalent complexes, and the VirA^{CDD}-VirFG^{NDD} and VirA^{ACP_{sb}-CDD}-VirFG^{NDD} covalent fusions and with a 2-fold symmetry restraint for the VirA^{ACP_{sb}-CDD}-VirFG^{NDD}-KS₆ fusion protein. (In the absence of such an imposed symmetry, the fit between the experimental and model SAXS curves was poorer (greater χ^2)). The results of several DAMMIN models were then compared using the program SUPCOMB⁴³ in order to determine a consensus model. Low-resolution models were then averaged using the DAMAVER and DAMFILT packages.^{44,45} The quality of the 3D modeling was determined using the discrepancy χ^2 , defined according to Konarev et al.⁴⁴ Values lying in the range of 0.9–1.1 are accepted to indicate a good fit between the models and the data. However, the calculation of χ^2 is inversely proportional to the measurement error, that is, the lower the error in the measurement, the higher the χ^2 . Because the detector at SOLEIL yields inherently low-error, high-quality data, the effect is to raise the determined χ^2 values above this standard range.⁴⁶ Nonetheless, visual inspection of the agreement between the theoretical curves calculated from the average molecular forms and the acquired SAXS data confirms the goodness of fit in all cases. Therefore, we decided not to rescale artificially the error on the measurement to give χ^2 of approximately 1, as has been suggested.⁴⁶

High-resolution structures or models were then superimposed into the *ab initio* protein shapes using SUPCOMB.⁴³ Figures displaying the protein structures were generated using PYMOL (Schrodinger, LLC), and the low-resolution bead models from DAMMIN were converted into mesh.

Molecular Modeling. A 3D structural model of the KS₆-linker region was generated by pairwise comparison of profile hidden Markov models using the HHpred server (<http://toolkit.tuebingen.mpg.de/hhpred>).⁴⁷ 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 KS₆-linker within the Protein Data Bank is KS₂ of PksJ (PDB: 2NA1) (35% sequence identity; 63% similarity).⁴⁸ This structure was then used to construct a model of the KS₆-linker using the associated MODELER software.⁴⁹ The KS₆ homodimer was reconstituted by superimposing two models onto the dimeric crystal structures of the PksJ KS.

Demonstration by SAXS That VirFG^{NDD} Undergoes a Conformational Change. *Ab initio* molecular forms for VirFG^{NDD} were calculated using DAMMIN.⁴² Twenty of the resulting low-resolution models (χ^2 of 2.321–2.489 vs the SAXS data) were compared using SUPCOMB,⁴³ revealing an extended shape that is not consistent with the conformation of VirFG^{NDD} adopted in the VirA^{CDD}-VirFG^{NDD} fusion. Indeed, calculation of theoretical scattering curves for VirFG^{NDD} alone using CRY SOL⁵⁰ based on the NMR structure of the VirA^{CDD}-VirFG^{NDD} fusion (all 20 structures within the ensemble were used) uniformly give poor fits to the

experimental data for scattering angles between 0.0005 and 0.3 Å⁻¹ ($\chi^2 = 6.61\text{--}10.39$). Furthermore, the Kratky plot⁵¹ computed from the experimental SAXS curve is consistent with an elongated shape/flexible particle. Taken together, these calculations clearly show that free VirFG^{NDD} and VirFG^{NDD} in complex with partner do not adopt the same structure, consistent with an induced conformational change.

Accession Numbers. Chemical shifts for VirA^{CDD} in complex with VirFG^{NDD} and for VirA^{CDD}–VirFG^{NDD} have been deposited in the Biological Magnetic Resonance Bank with accession codes 25721 and 25706, respectively. PDB coordinates for VirA^{CDD}–VirFG^{NDD} have been deposited under accession code 2NSD.

RESULTS

In Silico Analysis of Docking in *trans*-AT PKSs. In earlier work, we identified VirA^{CDD} as comprising the 50 residues C-terminal of the conserved ACP_{sb} domain of module 5 (Figure 1).²¹ 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 (Figure S1). Although BLAST⁵² analysis of this region returned no homologues within the PDB, the first 28 residues of VirFG^{NDD} show low sequence homology to VirA^{CDD} (28% identity⁵³) and to multiple Class 2 docking domains (e.g., 4–21% to the six reported docking domains of the curacin PKS;¹⁷ Figure S1). The C-terminal 42 amino acids, however, 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 PKSs, allowing putative^{CDD} and^{NDD} 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.^{6,15–17,54} Our analysis of 44 published *trans*-AT PKSs reveals that such interfaces are formed by a much greater diversity of domain pairs (Table S2). 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, ketoreductase (KR), ER, or C/heterocyclization (HC)) revealed regions with low but suggestive homology to VirA^{CDD} (10–53% identity; avg. 22%)⁵³ and VirFG^{NDD} (8–39% identity; avg. 23%),⁵³ respectively. All of the regions were predicted by secondary structure analysis (PsiPred²³ and PROMALS3D)²⁴ to contain two α helices or, more rarely, one long α helix (Figure S2).

The 27 identified junctions (Table S3) include the ACP/KS interfaces present in the anthracimycin, chivosazol, elansolid, enacyloxin, kirromycin, oxazolomycin, rhizoxin, and thailandamide systems (four of these junctions are of the specific type tandem ACP-ACP/KS as in the Vir system), the ACP/C or ACP/HC (PKS/NRPS) interfaces within the bacillaene, chivosazol, disorazol, kalimantacin/batumin, kirromycin, oxazolomycin, rhizopodin, and thailandamide assembly lines, an ACP/KR interface within the macrolactin system, and ACP/ER junctions in the bryostatin, etnangien, and myxovirescin machineries. *Streptomyces pristinaespiralis* also harbors a hybrid *trans*-AT PKS-NPRS that generates virginiamycin (though the

product is referred to as pristinamycin IIA), and as expected, sequences with homology to VirA^{CDD} and VirFG^{NDD} were present on the respective subunits SnaE2 and SnaE3.⁵⁵ Nonetheless, the^{NDD} 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 65 054); translation of this longer region gave a sequence incorporating a complete N-terminal docking domain (Figure S2). Conversely, the published N-terminus of TaiM from the thailandamide PKS⁵⁶ begins some 300 residues upstream of a Met and the first conserved α helix of the putative docking domain, again suggesting an incorrectly assigned start site. In contrast, the TaiE subunit from the same PKS also contains a long extension N-terminal to the docking domain, but its presence cannot apparently be explained by a misassigned start codon.

VirA^{CDD} and VirFG^{NDD} Serve as Docking Domains.

Having identified VirA^{CDD} and VirFG^{NDD} as representative of their respective families of C- and N-terminal docking domains, we aimed to investigate their interaction. For this, we expressed and purified (Figure S3) the following constructs in recombinant form from *E. coli* BL21(DE3) as His₆-SUMO-tagged proteins using plasmid pBG-102 (Vanderbilt University Center for Structural Biology): VirA^{CDD}, VirFG^{NDD}, VirA^{ACP_{sb}-CDD}, VirFG^{NDD}-linker (all 70 residues up to the conserved KS₆), VirFG^{NDD}-KS₆, and a protein (VirA^{CDD}–VirFG^{NDD}) in which the two docking domains were covalently fused via the same eight-amino-acid linker (G₃SG₃S) used previously to study the Class 2 docking domains of *cis*-AT PKSs¹⁷ (Supporting Information Methods). For the sequences of all constructs, see Table S4. To allow both efficient monitoring of VirA^{CDD} and VirFG^{NDD} by UV–vis during purification and dosing for binding studies by ITC, we added a tyrosine residue at the N- and C-terminal ends, respectively; subsequent structural analysis (*vide infra*) showed that these additions had no effect on the native docking interaction involving the opposite termini. Following cleavage of the His₆-SUMO-tag using human rhinovirus 3C (H3C) protease, the N-termini of all of the proteins also harbored a non-native GPGS sequence; this sequence is highly similar to and occupies the same position relative to VirFG^{NDD} as the G₃SG₃S sequence used to fuse the domains covalently, which we also show by NMR (*vide infra*) to have no effect on the docking interaction while in the case of VirA^{CDD}, it lies upstream of the added Y. In the case of VirFG^{NDD}-KS₆, we designed the expression construct to include 179 residues downstream of the KS₆ domain (encompassing consecutive regions showing homology to the so-called KS-AT and post-AT linkers observed in structures of KS-AT didomains from *cis*-AT PKSs)¹⁸ because this choice of boundary previously allowed us to obtain the KS of module 5 as a stable protein.²¹

As an initial demonstration that the docking domains interact, we carried out analytical gel filtration in the presence of VirA^{CDD}, VirFG^{NDD}, and a 1:1 mixture of the two domains (all at an initial concentration of 670 μ M). On their own, VirA^{CDD} and VirFG^{NDD} gave elution volumes of ca. 13.7 and 14.5 mL, respectively, whereas the 1:1 mixture gave a single peak with a new elution volume of 14 mL (Figure S4), providing the first evidence for a specific interaction between the two docking domains. Conversion into molecular weights yielded the following data: VirA^{CDD}, 11.8 kDa (monomer

weight 4.3 kDa); VirFG^{NDD} 7.4 kDa (monomer weight 4.1 kDa); and the VirA^{CDD}/VirFG^{NDD} complex, 9.9 kDa (heterodimeric complex weight, 8.4 kDa). These data suggested that both domains are oligomeric when isolated but recombine to form a 1:1 complex; in fact, VirFG^{NDD} is monomeric but elutes anomalously because of its extended conformation (*vide infra*). Measurement of the affinity of the interaction by ITC yielded a value of $5.8 \pm 0.2 \mu\text{M}$ (from three independent analyses, Figure S5), which is within the range measured for Class 1 and 2 docking domains ($4\text{--}20 \mu\text{M}$);^{16,17} a similar K_d ($4 \mu\text{M}$) was measured for the VirA ACP_{sb}-^{CDD}/VirFG^{NDD} pair (Figure S5). Inclusion of the 42-residue linker downstream of VirFG^{NDD} did not significantly alter the affinity ($K_d = 5.5 \mu\text{M}$ for the interaction between VirA^{CDD} and VirFG^{NDD}-linker, Figure S5), supporting our supposition that this region is not involved in docking.

Characterization of Docking between VirA^{CDD} and VirFG^{NDD} by NMR, SAXS, and Allied Methods. To investigate in detail the interaction between the two docking domain partners, we analyzed ¹⁵N-labeled VirA^{CDD} in isolation and in the presence of unlabeled (¹⁴N) VirFG^{NDD} so that in either case only signals arising from VirA^{CDD} would be visible in the obtained spectra. In the absence of its partner docking domain, the ¹⁵N-¹H HSQC spectrum of VirA^{CDD} ($100 \mu\text{M}$) exhibited only 15 of 37 expected peaks (Figure S6). In contrast, when a 2- to 4-fold excess of ¹⁴N VirFG^{NDD} was also present, 36 of 37 peaks were observed (Figure S6). Two explanations account for the fewer than expected number of peaks in the spectrum of unpartnered VirA^{CDD}: (i) an equilibrium between folded and unfolded states, on the intermediate NMR time-scale, or (ii) a mixed population of nonspecific homodimers (a state observed previously),²¹ which also undergo exchange on the intermediate time-scale. (The set of residues at the interfaces are those which disappear because they experience multiple chemical environments.) We strongly favor the second explanation because data obtained by ITC as the concentration of VirA^{CDD} was increased incrementally are consistent with a net monomer to dimer transition, with the protein dimerizing completely above $30 \mu\text{M}$ (Figure S5). Furthermore, molecular weight estimation by SAXS MoW following SAXS analysis of VirA^{CDD} (Figure S7 and Table S5) yielded a molecular weight of 8.4 kDa (this procedure typically produces measurement errors of $<10\%$;⁴⁰ homodimeric complex theoretical weight 8.6 kDa), and the result was confirmed by SEC-MALLS, which gave a molecular weight of $8.2 \pm 0.6 \text{ kDa}$. Using this information, we calculated average molecular forms for VirA^{CDD} in solution, with and without imposition of second-order symmetry. In all cases, the fits between the theoretical curves calculated from the forms and the SAXS data (for the q range 0.018 and 0.3 \AA^{-1}) were good, but the χ^2 values were lower in the absence of imposed symmetry ($1.732\text{--}2.543$ vs $1.889\text{--}3.28$, Figure S7). The overall heterogeneity of these shapes supports the idea that VirA^{CDD} in solution is best described as a population of interchanging, nonspecific dimers. To investigate whether VirA^{CDD} retains its structure even in the monomeric form, we analyzed it by CD at concentrations both above ($100 \mu\text{M}$) and below ($10 \mu\text{M}$) the estimated dimerization point. This experiment showed that although autoassociation did increase the overall secondary structure, VirA^{CDD} retained 73% of its α -helical content in monomeric form (Figure S8).

Although isolated VirA^{CDD} clearly homodimerizes, in the presence of excess VirFG^{NDD} it undergoes subunit exchange

to form a stable, heterodimeric docked complex, explaining the observation of almost all anticipated NMR peaks. This fact allowed us to fully assign ¹⁵N,¹³C-VirA^{CDD} within the context of this complex using a combination of 2D and 3D NMR experiments in order to determine its secondary structure (Figure S6). Consistent with the secondary structure predictions, bound VirA^{CDD} comprises two short α helices (residues Q6936–R6947 and I6953–L6960; numbering based on GenBank: BAF50727).

We next carried out these experiments in reverse, with ¹⁵N-labeled VirFG^{NDD} and unlabeled (¹⁴N) VirA^{CDD}. The ¹⁵N-¹H HSQC spectrum of isolated VirFG^{NDD} ($100 \mu\text{M}$) showed the expected number of peaks, but with poor chemical shift dispersion (Figure S6). Dynamical studies also showed it to be highly flexible (Figure S6). Taken together, these results indicated that VirFG^{NDD} alone (i.e., in the absence of its docking partner or a downstream catalytic domain) lacks a well-defined 3D structure and so could be an intrinsically disordered protein (IDP).⁵⁷ (For a discussion of the applicability of these findings to other *trans*-AT PKS, see the Appendix). In agreement with this, CD analysis yielded data consistent with only partial structuration (22% α helix, Figure S8), whereas analysis by ITC under the same conditions as for VirA^{CDD} provided no evidence for homodimerization (Figure S5). As further support for the IDP character of VirFG^{NDD}, addition of 2,2,2-trifluoroethanol (TFE) to the buffer significantly increased the α -helical content as measured by CD (Figure S8). 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.⁵⁸ We also analyzed VirFG^{NDD} by SAXS under the NMR conditions, yielding an R_g of $16 \pm 1 \text{ \AA}$, a D_{max} of 66 \AA , and M_w s of 5 kDa (SAXS MoW) and $4.2 \pm 0.4 \text{ Da}$ (SEC-MALLS), consistent with a monomer (Figure S7 and Table S5). The D_{max} value accords with an extended, albeit partially folded protein (predicted D_{max} for a fully extended protein of 40 residues = 160 \AA , Figure S7); this result also explains the anomalous behavior by analytical GF, in which VirFG^{NDD} appeared to elute as a dimer because of its extended shape. The form of a Kratky plot⁵¹ 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 α helix (Figure S7). However, despite extensive attempts, we were unable to obtain a recombinant protein in which VirFG^{NDD} was attached to the downstream KS domain (VirFG^{NDD}-KS₆) because the construct aggregated in solution. Thus, we cannot strictly rule out the possibility that VirFG^{NDD} is structured in its native context even in the absence of VirA^{CDD}, although clearly it shows multiple characteristics of an IDP.

In any case, addition of a 4-fold excess of ¹⁴N-VirA^{CDD} (ca. $400 \mu\text{M}$) provoked a number of spectral modifications, including significant changes in chemical shift and an overall increase in chemical shift dispersion (Figure S6). These types of spectral changes are consistent with a coupled process of binding/induced structuration of VirFG^{NDD} by VirA^{CDD}. To bolster the case that VirFG^{NDD} was gaining structure upon contact with VirA^{CDD}, we carried out differential CD analysis, acquiring spectra of VirFG^{NDD} and VirA^{CDD} alone and in complex (Figure S8). This analysis showed that the spectrum of the complex does not correspond to a simple average of the spectra of the two individual partners (α -Helical content increased from 40 to 52%).³⁵ VirA^{CDD} is already at least

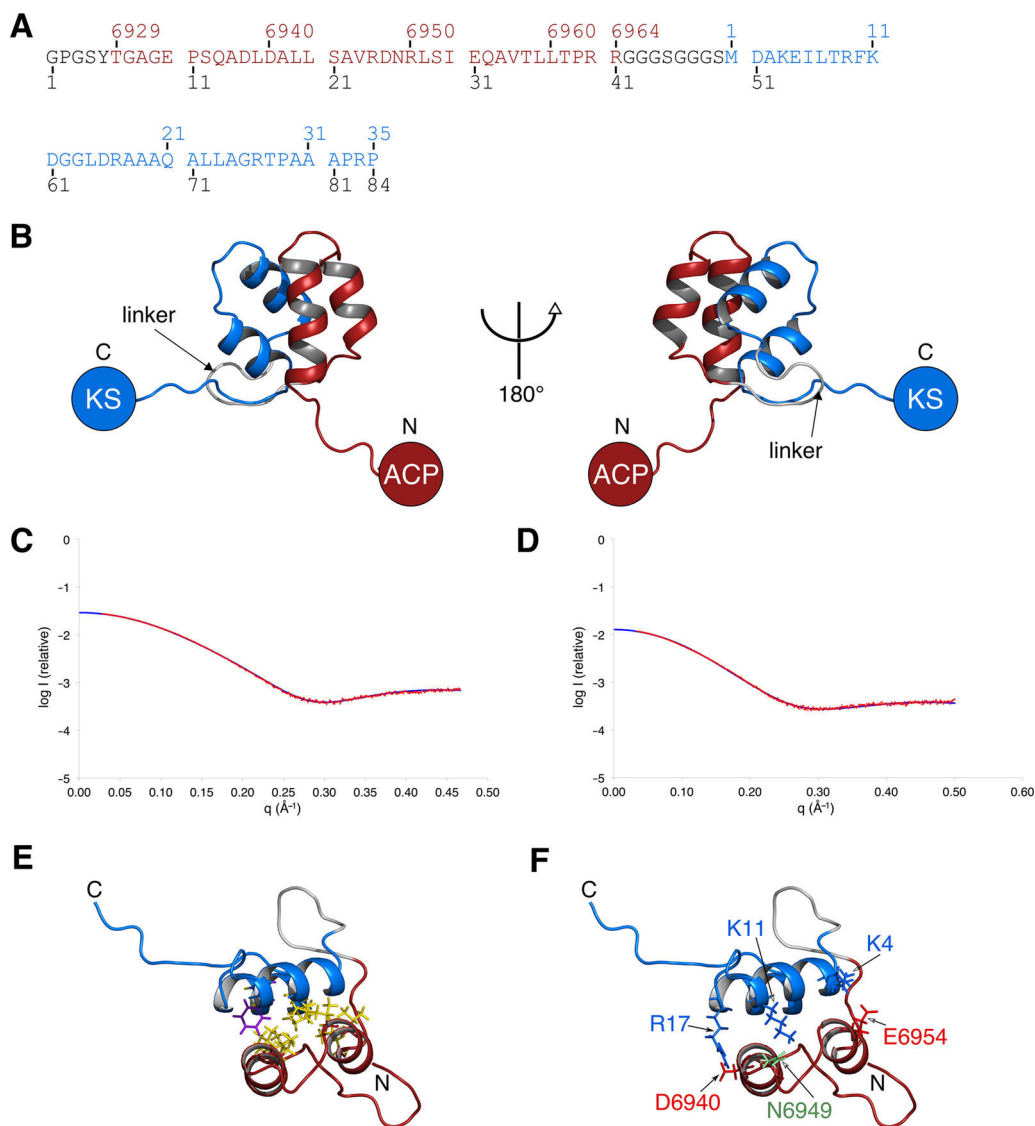


Figure 3. Analysis of docking between VirA^{CDD} and VirFG^{NDD}. (A) Sequence of the VirA^{CDD}-VirFG^{NDD} fusion construct, with VirA^{CDD} indicated in red, and VirFG^{NDD} indicated in blue. The N-terminal vector-derived residues and the added 8-residue linker are shown in black. The numbering above the sequence corresponds to that in GenBank, whereas that below is based on the construct itself. (B) NMR structure of the covalent docking domain complex VirA^{CDD}-VirFG^{NDD} color-coded as in A. (C and D) The strong correspondence between the SAXS curve for VirA^{CDD}-VirFG^{NDD} predicted from its NMR structure (CRY SOL⁵⁰) (blue) and those obtained (red) for (C) the noncovalent VirA^{CDD}/VirFG^{NDD} complex and (D) VirA^{CDD}-VirFG^{NDD} 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^{CDD}-VirFG^{NDD}. (F) Electrostatic interactions at the docking interface within VirA^{CDD}-VirFG^{NDD} that were targeted for site-directed mutagenesis.

partially structured at the concentration used in this experiment because of self-association; thus, we attribute the structuration seen during complex formation, at least in part, to induced folding of VirFG^{NDD}, although we cannot exclude that the secondary structure content of VirA^{CDD} also increases. An essentially identical ¹⁵N-¹H HSQC spectrum of ¹⁵N-labeled VirFG^{NDD} was obtained in the presence of VirA ACP_{5b}-CDD (Figure S6), consistent with the result from ITC showing that the presence of the ACP does not impact the docking domain interaction.

We followed up these observations by elucidating the complete structure of the docking domain complex, using the construct VirA^{CDD}-VirFG^{NDD} in which the two docking elements were covalently fused.¹⁷ (For the sequence of this protein, see Figure 3A.) Structural studies of ¹⁵N,¹³C-labeled VirA^{CDD}-VirFG^{NDD} (Figures 3 and S6) were carried out by

multidimensional heteronuclear NMR spectroscopy. Structure calculation was performed using a two-step procedure consisting of initial structure generation using CYANA³⁰ followed by restrained molecular dynamics refinement within Amber.^{33,34} The quality of the structures was assessed using PROCHECK-NMR⁵⁹ and the MolProbity⁶⁰ server (Table S6). The first important result is that the introduced linker between the two docking domains is unstructured in the complex (Figure 3B). Consistent with this, amide chemical shifts arising from this region are clustered together in the ¹⁵N-¹H HSQC spectrum, and dynamical studies show it to be highly mobile (Figure S6). Taken together, these data confirm that the linker acts simply as a flexible tether and does not contribute to the docking interface, supporting the idea that our fusion strategy has faithfully captured the native interaction.

Within the context of this fusion construct, VirA^{CDD} comprises two consecutive, short α helices as found for discrete VirA^{CDD} in complex with VirFG^{NDD}. Likewise, fused VirFG^{NDD} incorporates two short α helices (residues A3–K11 and R17–L24, numbered according to our revised sequence of VirFG,²¹ GenBank: KT308172), and its NMR spectrum superimposes on that of the domain in noncovalent complex with VirA^{CDD} (Figure S6). The introduced Y residues at the extremities of VirA^{CDD} and VirFG^{NDD} both lie within unstructured linkers at a significant distance from these key secondary structure elements (7 and 11 residues, respectively, for VirA^{CDD} and VirFG^{NDD}) and do not contact the docked complex, confirming that they have not affected the interaction. Together, the four α helices form a four α -helical bundle in which the two sets of α 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,^{15–17} 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.

Analysis by SAXS MoW and SEC-MALLS confirmed the heterodimeric nature of the covalent complex (calcd M_w 8.4 kDa; heterodimeric complex weight, 9 kDa via SAXS MoW; and 8.9 ± 0.9 kDa, via SEC-MALLS; Table S5). 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 VAST⁶¹ and DALI⁶² did not reveal a comparable topology for any known interprotein complexes, consistent with previous findings that IDPs characteristically form unusual interactions.⁶³ Independent evidence for the structure was obtained by SAXS because the curve predicted using CRY SOL⁵⁰ from the NMR structure of the VirA^{CDD}–VirFG^{NDD} fusion protein superimposed nearly perfectly with the SAXS data obtained on this construct ($\chi^2 = 2.90$; Figures 3C,D and S7 and Table S5). Furthermore, the SAXS curves predicted by CRY SOL for VirFG^{NDD} alone if it adopted the same structure as in the complex (i.e., those arising from all 20 structures within the NMR ensemble) clearly do not correspond to the experimental data obtained on the isolated protein ($\chi^2 = 6.61–10.39$), proving that VirFG^{NDD} undergoes a structural transition in the presence of its partner VirA^{CDD}. (See Figure S7 for a representative fit.) Within the complex, the interface is $\sim 730 \text{ \AA}^2$, an extensive surface area given the limited number of participating residues (17 amino acids within a 4 \AA radius). This too is consistent with VirFG^{NDD} being an IDP, as such regions are typically able to achieve a greater interface area/residue than comparable ordered proteins.⁶⁴ If two such complexes form simultaneously *in vivo* at the VirA/VirFG junction, then the total interaction surface area would be $\sim 1460 \text{ \AA}^2$, which is in between Class 1 ($\sim 1100 \text{ \AA}^2$) and Class 2 ($\sim 1800 \text{ \AA}^2$) docking domains and in accord with the affinity measured by ITC. The docking interface is hydrophobic, and comprises residues L6939, L6943, V6946, I6953, A6956, V6957, and L6960 of VirA^{CDD} and residues A3, L7, F10, A20, and L24 of VirFG^{NDD} (Figure 3E,F). The contact surface between the two VirA^{CDD} α helices includes L6942, A6945, L6951, A6956, and L6959, whereas that between the two VirFG^{NDD} α helices comprises I6, L15, A19, and L23. Inspection of the sequence alignments of the *trans*-AT PKS C- and N-terminal docking domains (Figure S2 and Table S7) shows that hydrophobic residues are relatively

well-conserved at the corresponding positions. Three electrostatic interactions are also present at the interface (residues (VirA^{CDD}/VirFG^{NDD}): D6940/R17, N6949/K11, and E6954/K4).

Analysis of Protein–Protein Interaction Determinants by Site-Directed Mutagenesis and ITC. Because 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 because these would also have an effect on measured K_d values. In fact, the CD experiments revealed that VirFG^{NDD} underwent architectural alterations when mutated, even gaining increasing secondary structure (Figure S8). This observation is again consistent with VirFG^{NDD} being an IDP and with its disordered character needing to be actively maintained against mutations that can increase structural content.⁶⁵ However, this behavior complicated interpretation of ITC data obtained using mutant VirFG^{NDD} because we could not clearly distinguish between mutations which disrupted an important interaction with VirA^{CDD} and those which simply impacted the ability of VirFG^{NDD} to undergo induced structuration. We therefore focused our attention on the VirA^{CDD} 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 because of poor solubility, but all other mutant proteins were obtained in acceptable yield. Far-UV CD analysis confirmed that their structures were not perturbed significantly relative to wild-type VirA^{CDD} (Figure S8).

Because of 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, Figure S5.) 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. The effect of changing the charge state of the residues was equally significant, with the affinity of VirFG^{NDD} binding to the VirA^{CDD} N6949K and E6954K mutants reduced to 37 ± 4 and $49 \pm 16 \mu\text{M}$, respectively. Even more dramatically, interaction between the D6940R mutant and VirFG^{NDD} was completely abolished. On the basis of these results, we decided to test the effect of a combined mutation, introducing D6940A and E6954A simultaneously into VirA^{CDD}. Analysis of the double mutant by CD showed that the structure was essentially unchanged, and the measured affinity to wild-type VirFG^{NDD} 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^{CDD}/VirFG^{NDD} complex. Inspection of the residue positions corresponding to D6940, N6949, and E6954 in the other docking domains of the putative family shows that nonpolar residues are also present in a number of systems and that even when the residues are charged they are often but not systematically matched by an opposite charge on the partner docking domain (Figure S2 and

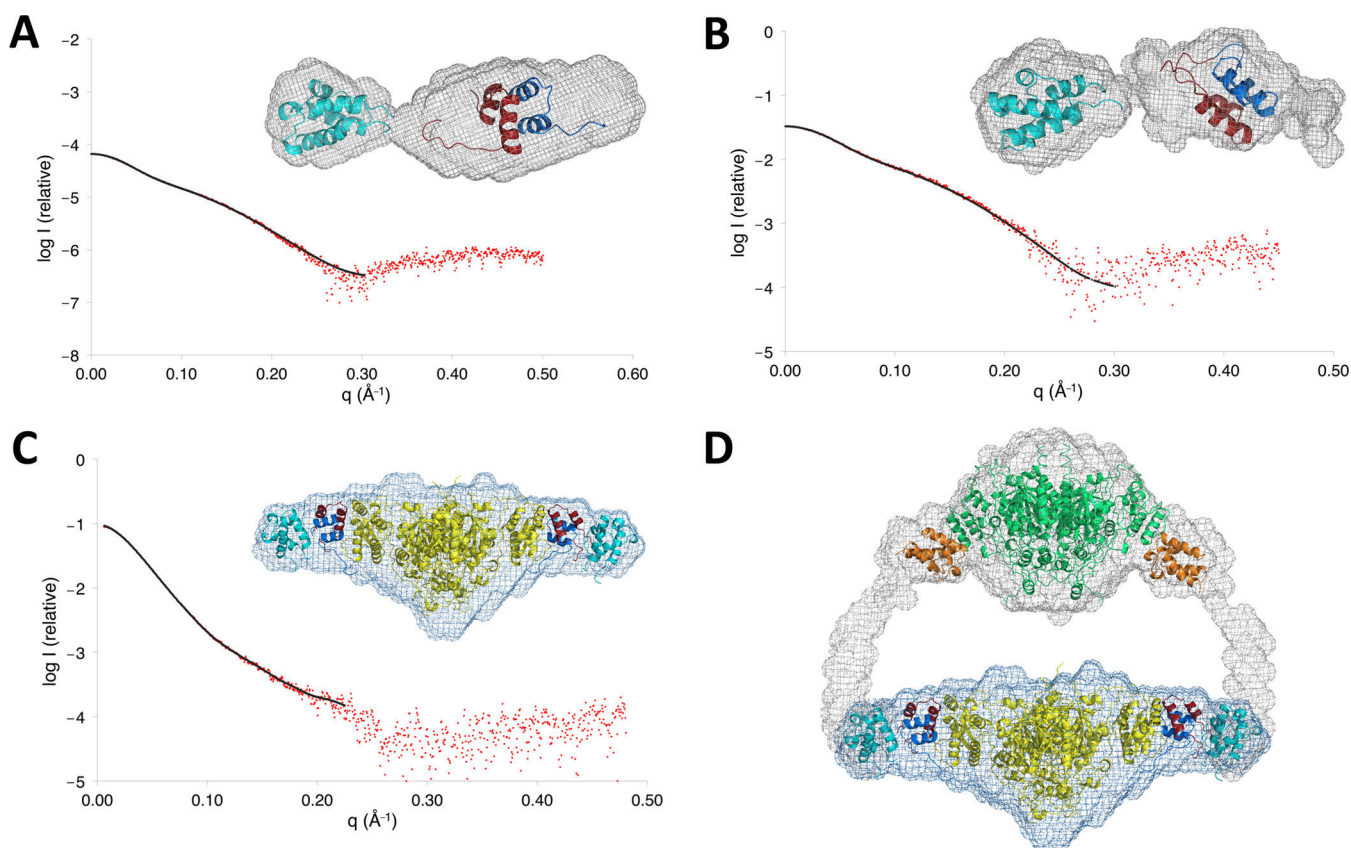


Figure 4. Analysis of the interface between subunits VirA and VirFG by SAXS. (A) Analysis by SAXS of the VirA ACP_{sb}-^{CDD}/VirFG ^{NDD} complex. Fit between the *ab initio* model computed with DAMMIN⁴² (solid black line) and the experimental SAXS data (red dots). This analysis yielded a molecular envelope different from that of VirA ACP_{sb}-^{CDD} alone²¹ because density was only observed for one ACP domain (light blue) and the docking domain complex (VirA ^{CDD} in red and VirFG ^{NDD} in dark blue). This result confirmed that docking occurs between a single copy each of the C- and N-terminal docking domains and that ACP_{sb} does not directly contact the complex of DDs. The presence of significant empty density likely reflects the fact that several species are present when the docking domains form a noncovalent complex. (B) Analysis of VirA ACP_{sb}-^{CDD}-VirFG ^{NDD} by SAXS, showing the fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots). Comparison with the average molecular form shown in A demonstrates that the covalent linkage of the two docking domains leads to a more homogeneous sample. (C) Analysis by SAXS of a covalent complex between VirA ACP_{sb}-^{CDD} and VirFG ^{NDD}-KS₆ (VirA ACP_{sb}-^{CDD}-VirFG ^{NDD}-KS₆). Fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots). A homology model of KS₆ (yellow) was placed into the average molecular envelope using SUPCOMB,⁴³ whereas localization of the NMR structure of ACP_{sb}²¹ 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 KS₅-ACP_{sa}-ACP_{sb} region of VirA module 5 (ref 21) (KS₅ in green, ACP_{sa} in orange) on the average molecular envelope obtained for VirA ACP_{sb}-^{CDD}-VirFG ^{NDD}-KS₆. This analysis suggests that the interaction between the subunits creates a single, closed reaction chamber within module 5. However, because the estimated distance between the ACP_{sb} and KS₆ active sites would not allow for direct chain transfer, conformational adjustment is likely to occur in the presence of polyketide intermediate on ACP_{sb}.⁷

Table S7); thus, it is not possible at present to discern a clear “code” that drives interaction specificity. 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 partner choice through effects on k_{on} .^{66,67} 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 (S922)); thus, they may also contribute to docking specificity through effects on k_{on} , k_{off} or both (Table S7).^{66,67}

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 3D structures, as is the case within the C- and N-terminal docking domains from the Chi,

Ozm, enacyloxin, and Tai systems (Table S3).⁶⁸ 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.⁵⁴ To address this issue, we aimed to identify by SAXS the locations of the flanking VirA ACP_{sb} and VirFG KS₆ domains relative to the docking domains. For this, we analyzed the following constructs: VirA ACP_{sb}-^{CDD}, VirA ACP_{sb}-^{CDD}/VirFG ^{NDD} complex, VirA ACP_{sb}-^{CDD}-VirFG ^{NDD} (a covalent fusion between ACP_{sb}-^{CDD} and VirFG ^{NDD}), and a fusion protein spanning from ACP_{sb} to KS₆ (VirA ACP_{sb}-^{CDD}-VirFG ^{NDD}-KS₆) (Figure S3); as mentioned previously, it was not possible to investigate noncovalent domain combinations involving VirFG ^{NDD}-KS₆ (i.e., VirA ^{CDD}/VirFG ^{NDD}-KS₆ complex and VirA ACP_{sb}-^{CDD}/VirFG ^{NDD}-KS₆ complex) because of aggregation. For the non-

covalent complex VirA ACP_{5b}-^CDD/VirFG^NDD, 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^CDD.

Previous analysis by SAXS of VirA ACP^{Sb}-^CDD²¹ yielded an average molecular envelope consistent with the homodimerization observed for VirA^CDD by NMR, ITC, and SAXS. In the presence of equimolar concentrations of its partner VirFG^NDD, however, a new average molecular envelope was obtained (Figures 4A and S7 and Table S5), in which density was observed only for a single ACP domain and the docking domain complex. Essentially identical data were acquired on VirA ACP_{5b}-^CDD–VirFG^NDD (Figures 4B and S7), strengthening the conclusion that joining the docking domains covalently does not alter how they interact. The molecular weight obtained in each case from SAXS MoW shows that only one copy of each docking domain is present in the complex (Table S5). 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_{5b} does not contact it directly. Analysis of VirA ACP_{5b}-^CDD–VirFG^NDD-KS₆ (Figure S7) gave an average molecule envelope ($\chi^2 = 1.77$) into which the homodimeric KS₆ domain could be confidently placed using SUPCOMB⁴³ (Figure 4C). It was not possible, however, to locate computationally the two ACP_{5b} domains and the two docking domain complexes because the scattering data are dominated by the signal from the KS₆. However, from the obtained molecular envelope it is clear that the 42-residue sequence separating VirFG^NDD from KS₆ must be largely structured because there is no electronic density which could correspond to an unstructured region of this size. Furthermore, making the reasonable assumption that the ACP_{5b} domains are located on the two extremities of the structure (Figure 4C) allowed us to superimpose the molecular envelope on that obtained previously for the portion of module 5 encompassing the KS₅, ACP_{5a}, and ACP_{5b} domains (KS₅–ACP_{5a}–ACP_{5b})²¹ (Figure 4D). This analysis shows that the two sets of ACP_{5b} domains superimpose nearly perfectly and therefore that the global form observed for VirA ACP_{5b}-^CDD–VirFG^NDD-KS₆ 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 ACP_{5b} to ACP_{5b} in module 5 alone = 279 Å²¹ vs 200 Å in KS₅–ACP_{5a}–ACP_{5b}), so as proposed earlier,²¹ interaction with VirFG induces closure of the module.

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).^{14–17} We have shown here that for an important subset of interpolypeptide junctions, analogous DDs operate in *trans*-AT PKSs. Of the 27 additional pairs of DDs identified by sequence/secondary structure analysis (Table S3), 10 are clearly predicted to exhibit properties (e.g., interaction propensity, extent of disorder, and lack of coiled-coil formation) in line with those of VirA^CDD and VirFG^NDD (Appendix), whereas the N-terminal docking domains at three other interfaces also appear similar to VirFG^NDD. For the remaining nine interfaces, however, the N-

terminal DD is predicted⁶⁹ to adopt a coiled-coil conformation. Comparative sequence analysis further suggests that this group of sequences may in fact be Class 2 type docking domains (Appendix), implying that multiple structural classes of docking elements also act in *trans*-AT PKS.

Overall, the recognition of four distinct structural classes of docking domains now permits a systematic comparison of docking strategies that give rise to both high specificity but relatively low affinity interactions (ca. 2–50 μM). All of the docking complexes incorporate contacts between α helices. Although architecturally distinct, the docking interfaces are largely hydrophobic and similar in size (1100–1800 Å²). 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 1b¹³).

Our work also suggests that 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 prokaryotes⁷⁰ (though relatively more common in high-GC, large-genome-size bacteria such as the Actinobacteria),⁷¹ and VirFG^NDD represents a strong candidate for the first IDP to be shown to participate in bacterial secondary metabolism. VirA^CDD 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 those of structured regions of the same proteins.⁶⁵ This is indeed the case for the Vir⁴ and Sna⁵⁵ systems, which both govern virginiamycin biosynthesis, because the corresponding subunits show much higher amino acid sequence identity than do 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).⁵³ 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 the structure of IDPs,⁶⁵ 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.

This 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 KS₆ domain of the next module. Indeed, the distance between the ACP_{5b} domains seems ideally adapted to accommodate the homodimeric KS₆ and two flanking docking domain complexes. In this way, docking between the two subunits in the context of a larger multienzyme complex (for example, incorporating the *trans*-AT and β-processing cassette) could give rise to a closed space containing a single reaction center in which reactive intermediates are protected. However, the estimated distance between ACP_{5b} and KS₆ active sites in this *apo* structure (ca. 80 Å) 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.⁶ 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_{sb} is correctly positioned to interact with KS₆. 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 functional intersubunit interfaces in *trans*-AT PKSs as long as the receiving KS has appropriate substrate specificity.⁵⁶

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b13372.

Appendix. (PDF)

Additional results, methods, and references. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*arnaud.gruez@univ-lorraine.fr

*benjamin.chagot@univ-lorraine.fr

*kira.weissman@univ-lorraine.fr

Author Contributions

[†]J.D. and T.A. contributed equally to this work

Funding

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 Lorraine Region (Bonus Qualité Recherche (BQR) grants to K.J.W. and B.C).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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), and Dr. Christophe Velours for help with the SEC-MALLS experiments. 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.

■ REFERENCES

- (1) Hertweck, C. *Angew. Chem., Int. Ed.* **2009**, *48*, 4688.
- (2) Katz, L. *Methods Enzymol.* **2009**, *459*, 113.
- (3) Piel, J. *Nat. Prod. Rep.* **2010**, *27*, 996.
- (4) Pulsawat, N.; Kitani, S.; Nihira, T. *Gene* **2007**, *393*, 31.
- (5) Weissman, K. J.; Müller, R. *ChemBioChem* **2008**, *9*, 826.
- (6) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 512.
- (7) Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 560.

- (8) Gokhale, R. S.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Science* **1999**, *284*, 482.
- (9) Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Biochemistry* **2001**, *40*, 2326.
- (10) Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. *Nat. Biotechnol.* **2005**, *23*, 1171.
- (11) Menzella, H. G.; Carney, J. R.; Santi, D. V. *Chem. Biol.* **2007**, *14*, 143.
- (12) Reeves, C. D.; Ward, S. L.; Revill, W. P.; Suzuki, H.; Marcus, M.; Petrakovsky, O. V.; Marquez, S.; Fu, H.; Dong, S. D.; Katz, L. *Chem. Biol.* **2004**, *11*, 1465.
- (13) Thattai, M.; Burak, Y.; Shraiman, B. I. *PLoS Comput. Biol.* **2007**, *3*, e186.
- (14) Richter, C. D.; Nietlispach, D.; Broadhurst, R. W.; Weissman, K. J. *Nat. Chem. Biol.* **2008**, *4*, 75.
- (15) Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003**, *10*, 723.
- (16) Buchholz, T. J.; Geders, T. W.; Bartley, F. E., 3rd; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *ACS Chem. Biol.* **2009**, *4*, 41.
- (17) Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. *Chem. Biol.* **2013**, *20*, 1340.
- (18) Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11124.
- (19) Tang, Y.; Chen, A. Y.; Kim, C.-Y.; Cane, D. E.; Khosla, C. *Chem. Biol.* **2007**, *14*, 931.
- (20) Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999**, *6*, 117.
- (21) Davison, J.; Dorival, J.; Rabeharindranto, H.; Mazon, H.; Chagot, B.; Gruez, A.; Weissman, K. J. *Chem. Sci.* **2014**, *5*, 3081.
- (22) Gu, L.; Eisman, E. B.; Dutta, S.; Franzmann, T. M.; Walter, S.; Gerwick, W. H.; Skiniotis, G.; Sherman, D. H. *Angew. Chem., Int. Ed.* **2011**, *50*, 2795.
- (23) McGuffin, L. J.; Bryson, K.; Jones, D. T. *Bioinformatics* **2000**, *16*, 404.
- (24) Pei, J.; Grishin, N. V. *Methods Mol. Biol.* **2014**, *1079*, 263.
- (25) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. *Nat. Struct. Biol.* **2002**, *9*, 522.
- (26) Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T. *Nat. Chem. Biol.* **2012**, *8*, 615.
- (27) *Molecular Cloning: A Laboratory Manual*; Sambrook, J., Fritsch, E. F., Maniatis, T., Eds.; Cold Spring Harbor Laboratory Press: New York, 1989.
- (28) Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. *Anal. Chem.* **2012**, *84*, 5066.
- (29) Cavanagh, J.; Fairbrother, W.; Palmer, A.; Skelton, N. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press Inc.: New York, 1996.
- (30) Güntert, P. Automated NMR Structure Calculation with CYANA. In *Protein NMR Techniques*; Downing, A. K., Ed.; Methods in Molecular Biology series; Humana Press: Totowa, NJ, 2004; Vol. 278, pp 353–378.
- (31) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209.
- (32) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213.
- (33) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; Legrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerlin, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 14*; University of California: San Francisco, CA, 2014.
- (34) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

- (35) Böhm, G.; Muhr, R.; Jaenicke, R. *Protein Eng., Des. Sel.* **1992**, *5*, 191.
- (36) David, G.; Pérez, J. J. *Appl. Crystallogr.* **2009**, *42*, 892.
- (37) Pieper, R.; Ebert-Khosla, S.; Cane, D.; Khosla, C. *Biochemistry* **1996**, *35*, 2054.
- (38) Bycroft, M.; Weissman, K. J.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **2000**, *267*, 520.
- (39) Svergun, D. J. *Appl. Crystallogr.* **1992**, *25*, 495.
- (40) Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F. *J. Appl. Crystallogr.* **2010**, *43*, 101.
- (41) Knobloch, J. E.; Shaklee, P. N. *Anal. Biochem.* **1997**, *245*, 231.
- (42) Svergun, D. I. *Biophys. J.* **1999**, *76*, 2879.
- (43) Kozin, M.; Svergun, D. J. *Appl. Crystallogr.* **2001**, *34*, 33.
- (44) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *J. Appl. Crystallogr.* **2003**, *36*, 1277.
- (45) Volkov, V.; Svergun, D. J. *Appl. Crystallogr.* **2003**, *36*, 860.
- (46) Svergun, D. I.; Koch, M. H. J.; Timmins, P. A.; May, R. P. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. OUP: Oxford, U.K., 2013.
- (47) Söding, J.; Biegert, A.; Lupas, A. N. *Nucleic Acids Res.* **2005**, *33*, W244.
- (48) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-Clay, A. T. *Structure* **2014**, *22*, 444.
- (49) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusdhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. In *Current Protocols in Protein Science*; Wiley: Hoboken, NJ, 2007; Chapter 2, Unit 2.9.
- (50) Svergun, D.; Barberato, C.; Koch, M. J. *Appl. Crystallogr.* **1995**, *28*, 768.
- (51) Receveur-Bréchet, V.; Durand, D. *Curr. Protein Pept. Sci.* **2012**, *13*, 55.
- (52) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403.
- (53) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Mol. Syst. Biol.* **2011**, *7*, 539.
- (54) Wu, N.; Cane, D. E.; Khosla, C. *Biochemistry* **2002**, *41*, 5056.
- (55) Mast, Y.; Weber, T.; Gözl, M.; Ort-Winklbauer, R.; Gondran, A.; Wohlleben, W.; Schinko, E. *Microb. Biotechnol.* **2011**, *4*, 192.
- (56) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008**, *26*, 225.
- (57) Sibille, N.; Bernadó, P. *Biochem. Soc. Trans.* **2012**, *40*, 955.
- (58) Bourhis, J.-M.; Receveur-Bréchet, V.; Oglesbee, M.; Zhang, X.; Buccellato, M.; Darbon, H.; Canard, B.; Finet, S.; Longhi, S. *Protein Sci.* **2005**, *14*, 1975.
- (59) Laskowski, R. A.; Rullmann, J. A. C.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477.
- (60) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12.
- (61) Bailey, T. L.; Gribskov, M. *Bioinformatics* **1998**, *14*, 48.
- (62) Holm, L.; Rosenström, P. *Nucleic Acids Res.* **2010**, *38*, W545.
- (63) Uversky, V. N. *Protein Sci.* **2013**, *22*, 693.
- (64) Liu, Z.; Huang, Y. *Protein Sci.* **2014**, *23*, 539.
- (65) Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biochem.* **2014**, *83*, 553.
- (66) Selzer, T.; Schreiber, G. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 190.
- (67) Schreiber, G.; Haran, G.; Zhou, H.-X. *Chem. Rev.* **2009**, *109*, 839.
- (68) Schreiber, G.; Keating, A. E. *Curr. Opin. Struct. Biol.* **2011**, *21*, 50.
- (69) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.
- (70) Williamson, M. P.; Potts, J. R. *Biochem. Soc. Trans.* **2012**, *40*, 945.
- (71) Pavlović-Lažetić, G. M.; Mitić, N. S.; Kovačević, J. J.; Obradović, Z.; Malkov, S. N.; Beljanski, M. V. *BMC Bioinf.* **2011**, *12*, 66.

C

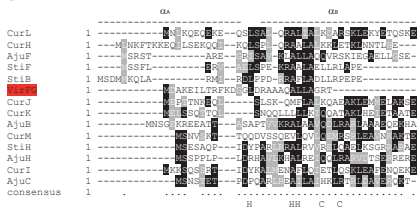


Figure S1. Comparative sequence analysis of VirFG^{DD} and Class 2 docking domains. (A) Sequence of subunit VirFG upstream of the conserved KS domain. The first 28 residues (in red) have been designated as VirFG^{DD}. (B) Multiple alignment using ClustalOmega¹ of the C-terminal regions of PKS subunits from *trans*-AT PKSs (in red), with those of *cis*-AT PKSs (black, except for those of subunit CurA and the closely-related JamE (in blue)) (analysis as published in ref. 3). The sequences (with the exception of VirFG^{DD}) from the *trans*-AT PKSs lie downstream of the conserved sequences of single (Tai) or tandem acyl carrier protein (ACP) domains associated with β-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 α-helices in the solved structure of the CurG and CurC C-terminal docking domains are indicated, as are the hydrophobic (H) and charged (C) residues which contribute to the docking interface.² (C) Multiple alignment using ClustalOmega¹ of the N-terminal regions of the partner *cis*-AT PKS subunits to those shown in (B), alongside that of VirFG (in red). The positions of the two α-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.² Key: Cur, curacin (*cis*-AT PKS);⁴ Rhi, rhizoxin (*trans*-AT);⁵ Mmp, mupirocin (*trans*-AT);⁶ Jam, jamaicamide (*cis*-AT);⁷ Sti, stigmatellin (*cis*-AT);⁸ Tai, thailandamide (*trans*-AT);⁹ Ta, myxovirescin (antibiotic Ta; *trans*-AT);¹⁰ Aju, ajudaloz (*cis*-AT);¹¹ Ela, elansolid (*trans*-AT);¹² Vir, virginiamycin (*trans*-AT);¹³ Sna, pristinaamycin (*trans*-AT).¹⁴

Table S2. 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		Atc, 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, Lum, Mis, Mln (2), Sor (3), Tst
KS*	DH	✓	Bae (2), Bat, Bon, Cal, Dif, Lum, Mis, Rhi, Tai, Tto
KS	DH*	✓	Cor, Dif
KS	C-MT	✓	Ta, Tai
KS	KR	✓	Ela, Lac, Lkc (2), Lum, Mgs, Mln, Riz
KS	KR*	✓	592
KS	KS	✓	Ooc
KS	ECH	✓	Cal
FH	ACP	✓	Ooc
KR	C-MT	✓	Bae, Bon, Cal, Dif (2), Ela, Etn, Ooc, Rhi, Sor (3)
KR	ACP	✓	Bon, Etn
DH	KS	✓	Lkc
DH	ACP	✓	Atc, Lac, Mgs, Ped
DH	KR	✓	Cto (2), Dis, Lkc (2), Lum, Mis, Mln (2), Tst, Tto
DH*	KR	✓	Mln
PS	KR	✓	Sor
C	A	✓	Alb
HC	A	✓	Cal, Rhi
C	GNAT	✓	Bry
C	C	✓	FR9, Tst
PCP	N-MT	✓	Ttt
O-MT	ACP	✓	Ozm, Nsp, Sna, Tdp(2)
C-MT	ACP	✓	FR9, Riz
Ox	KR	✓	FR9, Tst
E	C	✓	Tdp

*The full Vir-encoding region has not been published.¹³ But based on comparison with the highly-homologous *sna* cluster from *S. pristinaespiralis* which gives rise to the same molecule,¹⁴ 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, Fkfh-like domain; ACP, acyl carrier protein; C, condensation; HC, heterocyclization; ER, enoyl reductase; PS, pyrene 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:

Atc, anthracimycin;^{25,16} Chi, chivosazol;¹⁷ Cor, coralopyronin;¹⁸ Ela, elansolid;¹² FR9, FR901464;¹⁹ Kir, kirromycin;²⁰ Mmp, mupirocin;⁵ Ooc, oocycin;²¹ Ozm, oxazolomycin;²² Rhi, rhizoxin;⁵ Sna, pristinaamycin;¹⁴ Tai, thailandamide;⁹ Tdp, thailandepsin;²³ Tmp, thiomarolin;²⁴ Vir, virginiamycin;¹³ 592, enacyloxin;²⁶ Cal, calyculin;²⁶ Nsp, nosperin;²⁷ Onn, onnamide;²⁸ Ped, pederin;²⁹ Psy, psymberin;³⁰ Bry, bryostatin;³¹ Etn, etnamien;³² Ta, myxovirescin (antibiotic Ta);³³ Bae, bacillane;³³ Bat, kalmantacin/batumin;³⁴ Dis, disorazol;^{35,36} Riz, rhizopodin;³⁷ Dif, diffidin;³⁸ Lnm, leinamycin;³⁹ Lum, lumaialide;⁴⁰ Mis, misknolide;⁴⁰ Mln, macrolactin;⁴¹ Sor, sorangicin;⁴² Tst, thalastatin;⁴³ Bon, bongreic acid;⁴⁴ Tto, tolytoxin;⁴⁵ Lac, lactimidomycin;⁴⁵ Lkc, lankacin;⁴⁶ Mgs, mgrastatin;⁴⁷ Cto, chlorotoni;⁴⁸ Alb, albidin;⁴⁸ Ttt, thalassospiramide.⁴⁹

S5

S6

A



S7

S8



Figure S2. Identification of putative families of docking domains in trans-AT PKSs. (A) Secondary structure/sequence-based alignment by PROMALS3D⁵⁰ of VirA^C DD alongside putative C-terminal docking domains from trans-AT PKSs operating at junctions between ACP various partner domains (the sequences include all amino acids downstream from the conserved ends of the ACP domains). Predicted α-helices are indicated in red (those present in OzmN, KirAll and SnaE2 were predicted by independent analysis using PsiPred v. 3.2), and the extents of helices α1 and α2 in VirA^C DD as determined by NMR are shown above the alignments. Key to systems: Chi, chivosazol; Atc, anthracimycin; Kir, kirromycin; Ozm, oxazolomycin; Dis, disorazol; Tai, thailandamide; Vir, virginiamycin; Sna, pristiniomycin; 592, enacyloxin; Ta-1, myxovirescin; Rhi, rhizoxin; Riz, rhizopodin; Etn, entangien; Bae, bacillaene; Min, macrolactin; Bry, bryostatin; Bat, kalmintacin/batumin; Ela, elansolid. (B) Secondary structure/sequence-based alignment by PROMALS3D⁵⁰ of VirG^D DD alongside the putative partner docking domains of those shown in (A) (the sequences include all amino acids upstream from the conserved starts of the downstream catalytic domains). Predicted α-helices are indicated in red (those present in VirFG were predicted by independent analysis using PsiPred v. 3.2), and the extents of helices α1 and α2 in VirG^D DD as determined by NMR are shown above the alignments. In all cases, the downstream domain is indicated: ER, enoyl reductase; C, condensation; HC, heterocyclization; KS, ketosynthase; KR, ketoreductase. Key to color coding of interface residues: green = residue at hydrophobic interface in VirA^C DD-VirG^D DD complex structure; blue = intradomain hydrophobic interaction residue; purple = residue participating in an electrostatic interaction at the docking interface.

Table S3. Putative members of a new docking domain family in trans-AT PKS.

C-terminal domain	N-terminal domain	Systems/interfaces where present ^a
ACP	KS	AtcE/AtcF; ChiB/ChiC; ChiE/ChiF; ElaK/ElaO (ACP-ACP/KS); KirAV/KirAW; OzmN/OzmH; OzmI/OzmJ; RhiC/RhiD (ACP-ACP/KS); SnaE2/SnaE3 (ACP-ACP/KS); TaiE/TaiK (ACP-ACP/KS); TaiL/TaiM (ACP-ACP/KS); VirA/VirFG (ACP-ACP/KS); 5923/5922; 5925/5924
ACP	C	BaeM/BaeN; Bat1/Bat2; KirAll/KirAllI; OzmK/OzmL; TaiD/TaiE
ACP	HC	ChiC/ChiD; DisB/DisC; RizC/RizD
ACP	KR	MinD/MinE
ACP	ER	BryA/BryB; BryB/BryC; EtnH/EtnI; Ta-1/TaO

^aThe host organisms of these 19 systems include six Streptomyces strains (Streptomyces strains CNH365/T676 (anthracimycin, Atc), Streptomyces collinus Tu 365 (kirromycin, Kir), Streptomyces pristinaespiralis (pristinamycin, Sna), Streptomyces virginiae (virginiamycin, Vir) and Streptomyces albus JA43453 (oxazolomycin, Ozm)), five myxobacterial strains (Sorangium cellulosum So ce56 (chivosazol, Chi and entangien, Etn), Chitinophaga pinensis DSM2588 (Elansolid, Ela), Sorangium cellulosum So ce12 (disorazol, Dis), Stigmatella aurantica Sg a15 (rhizopodin, Riz) and Myxococcus xanthus DK1622 (myxovirescin, Tai)) three Burkholderia strains (Burkholderia rhizoxinica (Rhizoxin, Rhi), Burkholderia thailandensis E264 (thailandamide, Tai) and Burkholderia ambifaria AMMD (enacyloxin, 592)), two Bacillus strains (Bacillus subtilis 168 (bacillaene, Bae) and Bacillus amyloqueliciens FZB42 (macrolactin, Min)), and one Pseudomonas (Pseudomonas fluorescens BCCM_109359 (kalmintacin/batumin, Bat)). Thus, these docking domains show a limited phylogenetic distribution, but nonetheless occur within multiple phyla (actino- and proteobacteria and firmicutes). Interfaces derived from the same system are indicated by matching colors (i.e. red for enacyloxin). The interfaces in italics may actually incorporate Class 2 docking domains, as characterized from cis-AT PKS (see Appendix for further details of the analysis).

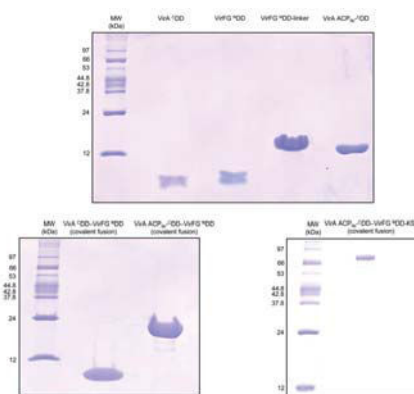


Figure S3. SDS-PAGE analysis of protein preparations used in this study. VirA^C DD (calc'd: 4.3 kDa); VirG^D DD (calc'd: 4.1 kDa); VirFG^D DD-linker (calc'd: 7.2 kDa); VirA^C DD-VirG^D DD (calc'd: 14.3 kDa); VirA^C DD-VirFG^D DD (covalent fusion) (calc'd: 8.4 kDa); VirA^C DD-VirFG^D DD (covalent fusion) (calc'd: 18.4 kDa); VirA^C DD-VirFG^D DD-VirG^D DD-VirG^D DD (covalent fusion) (calc'd: 85.8 kDa). The molecular weights of the markers are indicated.

Table S4. Sequences of constructs investigated in this study

Name of construct	Sequence ^a
VirA ^Δ DD	GP GSYTGAGEPS QADLDALLSAVR NRNLS IEQAVTLLT FRR
VirFG ^Δ DD	GP GSMDAKEI ILTRFKD GGLDRAAAQ ALLAGRT PAAAPRFY
VirFG ^Δ DD-linker	GP GSMDAKEI ILTRFKD GGLDRAAAQ ALLAGRT PAAAPRFSEPAAPTRPAVPAVPAVEPAA GTTVAEGTAGRF
VirA ACP ₉₅ ^Δ DD	GP GSVAVDPAPVARALREELARTLYCEPGDIDDEASNTLGLDSILGVEFVAVNQTYGL DEKAGILYDHFSLAALS SHVAGRAAPV PAAAAEPALPLPV
VirFG ^Δ DD-KS ₅	GP GSMDAKEI ILTRFKD GGLDRAAAQ ALLAGRT PAAAPRFSEPAAPTRPAVPAVPAVEPAA GTTVAEGTAGRFVAVIYGSAREFGAADDFWQRILDGDDLVTEVPERWRTTEFYDP DPAEGRSVRWGAYVADDFRDAFFRMFREAELDPAQLQEAARALEHGRDARS LAGTRCGVAGVMINDYQDLVERESPVKLPQVMQGNNSILAARIAYHLDLKGPVAVTDT ACSSLTALHLACQSLWLGEDTLTVGGVTLVTELEPHVMSAAGMLSPNGRCRPFDAAD GIVPEGCAVVVKPLSKALADGDPVHAVIRASGLNQDRGPGQGITAPARSQALVVRTLL QRFVAVDPAGIDYVECHGTOTPLGDPLEVTALNEAFAGAGLAPASVPIGVSVGNIGHTSAA AGLALGLKAGVVRTGLVPSLHYARANPQIPFDQGFVAGERRELGPDGGRPRATVVS SPGSGTNAVYVVEQAEQARPAAGDDGAPVLPISGPARADAPAAHADLARWLKRGPGS DASPADVAHTLAVARTHTYRSALLVSGRDELLESLDLAGSADPRRTDTPADAAAPDAA VRRQAALLARLVERAGENGPVLEALAKLYTQGHVDWAASPPARHRLSLPAYFPFAP HRHYVER
VirA ^Δ DD-VirFG ^Δ DD	GP GSYTGAGEPS QADLDALLSAVR NRNLS IEQAVTLLT PRR GGSGGGSMDAKE ILTRFKD GGLDRAAAQ ALLAGRT PAAAPRF
VirA ACP ₉₅ ^Δ DD-VirFG ^Δ DD	GP GSVAVDPAPVARALREELARTLYCEPGDIDDEASNTLGLDSILGVEFVAVNQTYGL DEKAGILYDHFSLAALS SHVAGRAAPV PAAAAEPALPLPV TGAGEPSQADLDALLSAVR NRNLS IEQAVTLLT PRR GGSGGGSMDAKE ILTRFKDGGLDRAAAQ ALLAGRT PAAAPRF
VirA ACP ₉₅ ^Δ DD-VirFG ^Δ DD-KS ₅	GP GSVAVDPAPVARALREELARTLYCEPGDIDDEASNTLGLDSILGVEFVAVNQTYGL DEKAGILYDHFSLAALS SHVAGRAAPV PAAAAEPALPLPV TGAGEPSQADLDALLSAVR NRNLS IEQAVTLLT PRR GGSGGGSMDAKE ILTRFKDGGLDRAAAQ ALLAGRT PAAAPRFSE PAAAPTRPAVPAVPAVEPAA GTVAAGTAGRPEP VAVIYGSAREFGAADDFWQRILDG DDLVTEVPERWRTTEFYDPDPAEGRSVRWGAYVADDFRDAFFRMFREAELDPAQLQEAARALEHGRDARS LAGTRCGVAGVMINDYQDLVERESPVKLPQVMQGNNSILAARIAYHLDLKGPVAVTDT ACSSLTALHLACQSLWLGEDTLTVGGVTLVTELEPHVMSAAGMLSPNGRCRPFDAAD GIVPEGCAVVVKPLSKALADGDPVHAVIRASGLNQDRGPGQGITAPARSQALVVRTLL QRFVAVDPAGIDYVECHGTOTPLGDPLEVTALNEAFAGAGLAPASVPIGVSVGNIGHTSAA AGLALGLKAGVVRTGLVPSLHYARANPQIPFDQGFVAGERRELGPDGGRPRATVVS SPGSGTNAVYVVEQAEQARPAAGDDGAPVLPISGPARADAPAAHADLARWLKRGPGS DASPADVAHTLAVARTHTYRSALLVSGRDELLESLDLAGSADPRRTDTPADAAAPDAA VRRQAALLARLVERAGENGPVLEALAKLYTQGHVDWAASPPARHRLSLPAYFPFAP HRHYVER

^aRed = sequence introduced from the plasmid; green = added Tyr; blue = linker used to fuse the docking domains. The α-helices present in VirFG^ΔDD and VirA^ΔDD (according to the NMR structure of the docked complex) are indicated in bold.

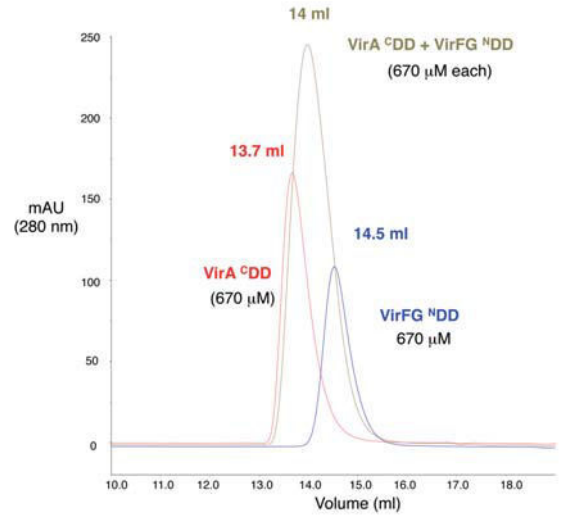
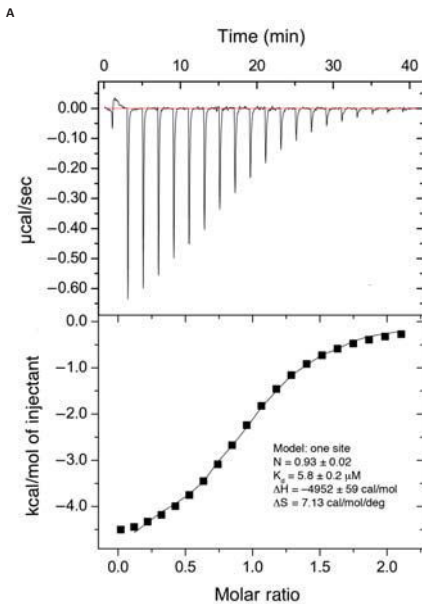


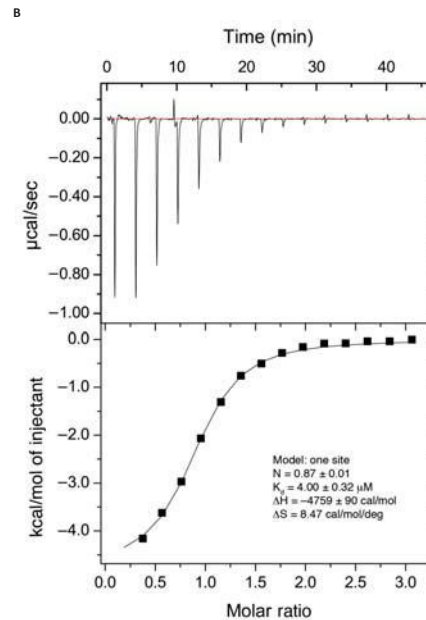
Figure S4. Analysis by gel filtration of the interaction between VirA^ΔDD and VirFG^ΔDD. The individual docking domains (both at an initial concentration of 670 μM) exhibit elution volumes of 13.7 ml and 14.5 ml, respectively. When combined at the same concentrations, a single new peak is observed at an intermediate elution volume (14 ml).

S13

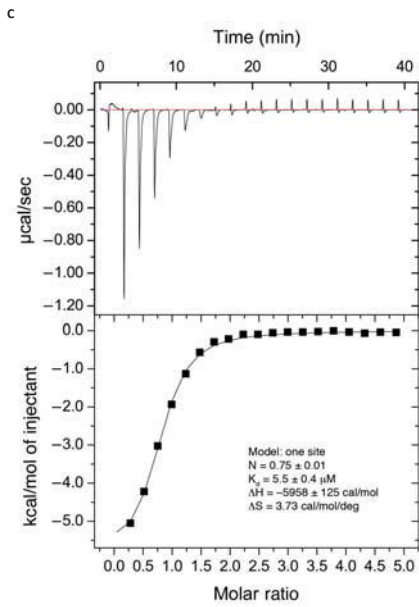
S14



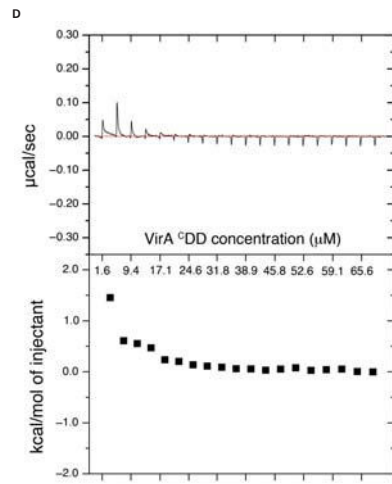
S15



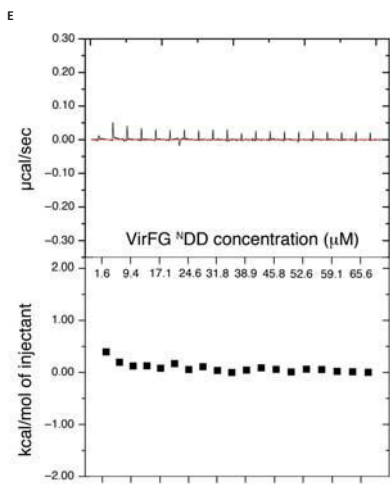
S16



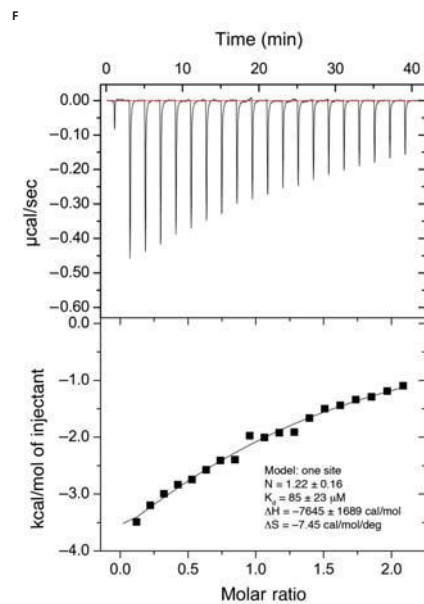
S17



S18

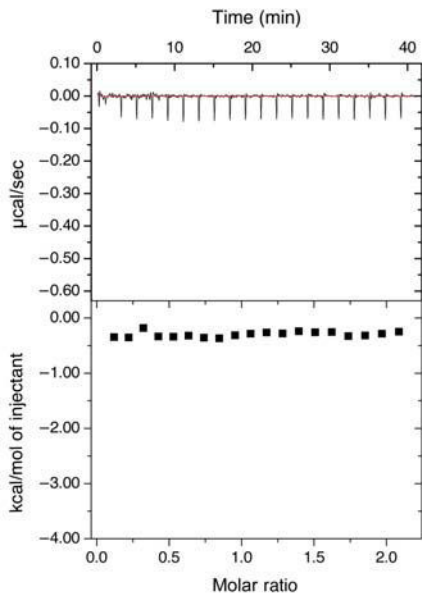


S19



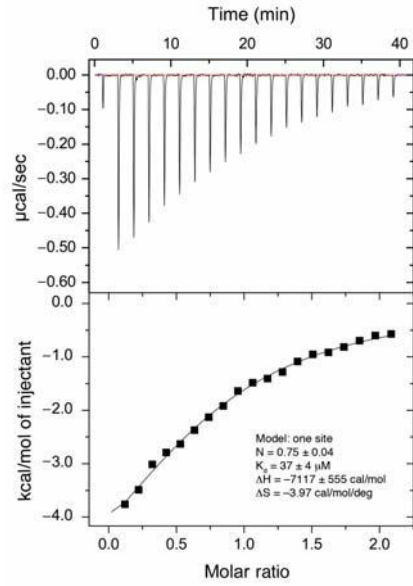
S20

G



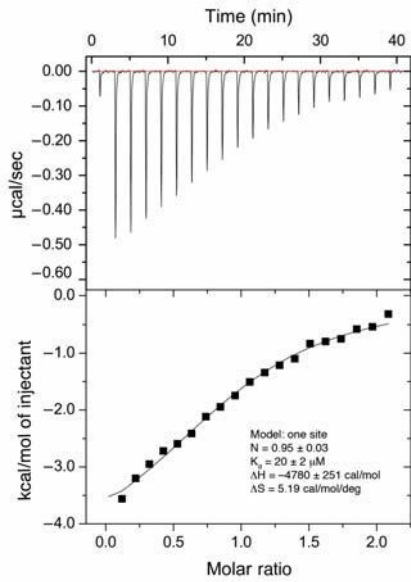
S21

H



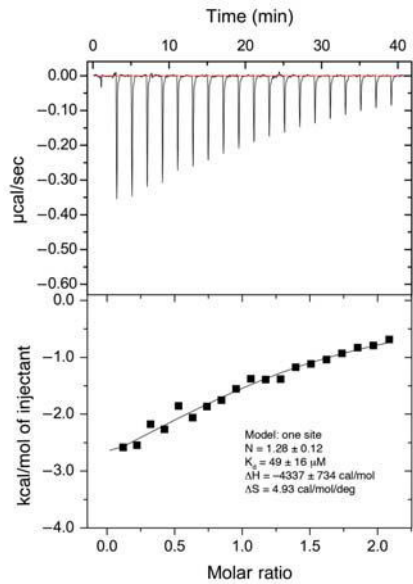
S22

I



S23

J



S24

K

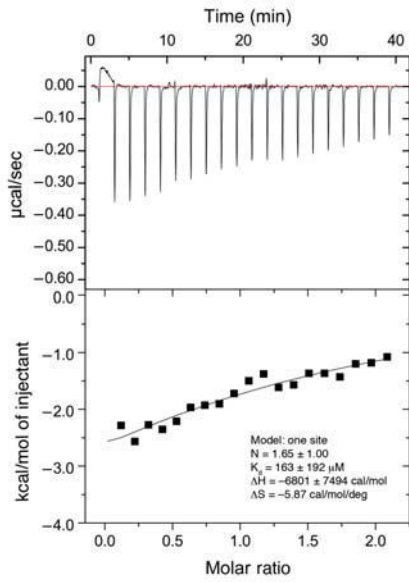
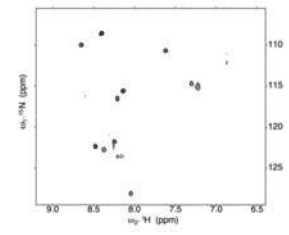


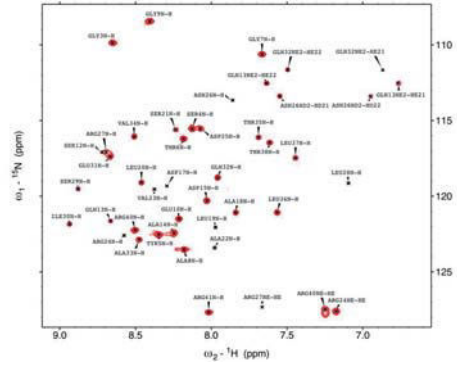
Figure S5. Determination by isothermal titration calorimetry of docking affinity. (A) Analysis of binding between wild type VirA^ΔDD and VirFG^ΔDD. (One of three measurements is shown, but the K_d value represents the average of three independent experiments). (B) Analysis of binding between wild type VirA^ΔACP₅₂-^ΔDD and VirFG^ΔDD. (C) Analysis of binding between wild type VirA^ΔDD and VirFG^ΔDD-linker. (D) Dilution analysis of VirA^ΔDD. No further change in heat signal is observed after VirA^ΔDD reaches a concentration of 30 μ M, indicating that it is completely homodimeric above this concentration. (E) Dilution analysis of VirFG^ΔDD. No evidence for homodimerization is observed across the full range of concentrations (to 78 μ M). (F) Analysis of binding between VirA^ΔDD D6940A and wild type VirFG^ΔDD. (G) Analysis of binding between VirA^ΔDD D6940R and wild type VirFG^ΔDD. (H) Analysis of binding between VirA^ΔDD N6949K and wild type VirFG^ΔDD. (I) Analysis of binding between VirA^ΔDD E6954A and wild type VirFG^ΔDD. (J) Analysis of binding between VirA^ΔDD E6954K and wild type VirFG^ΔDD. (K) Analysis of binding between VirA^ΔDD D6940A + E6954A and wild type VirFG^ΔDD. For (B)-(K), the indicated error is that on the individual measurement.

S25

A

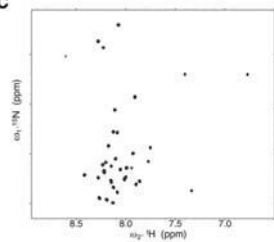


B

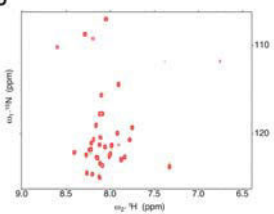


S26

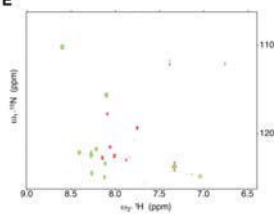
C



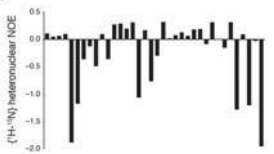
D



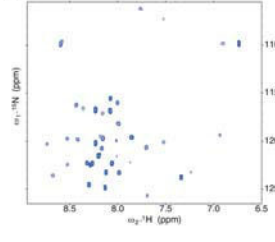
E



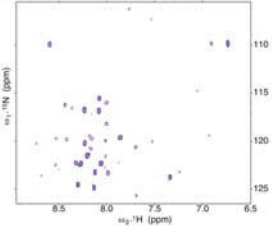
F



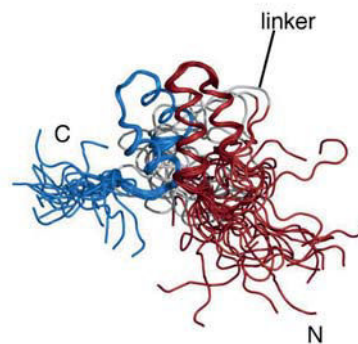
G



H



I



S27

S28

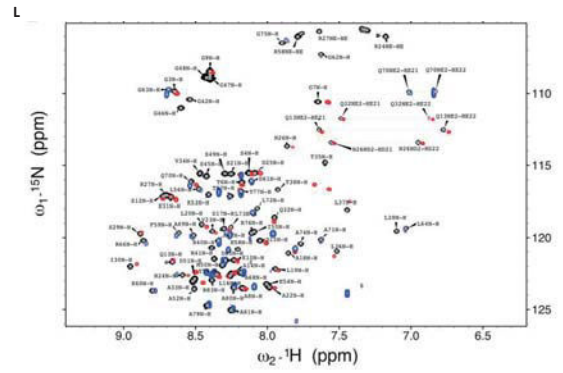
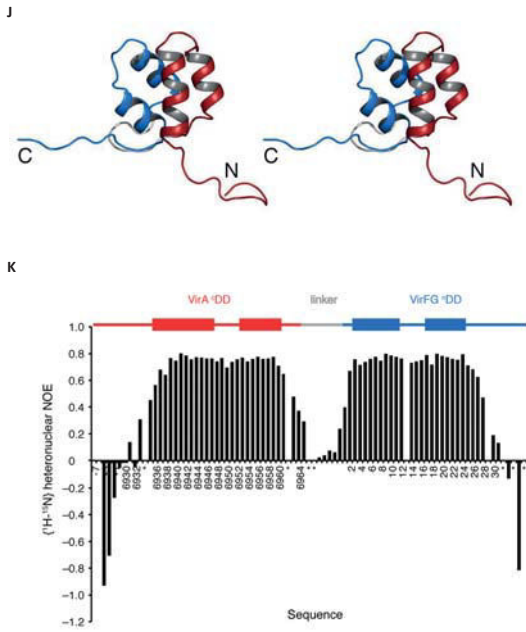


Figure S6. NMR studies of VirA $^{\Delta}$ DD and VirFG $^{\Delta}$ DD. (A) Unassigned ^1H , ^{15}N -HSQC spectrum of ^{15}N VirA $^{\Delta}$ DD alone. (B) Assigned ^1H , ^{15}N -HSQC spectrum of ^{15}N VirA $^{\Delta}$ DD in complex with ^{14}N VirFG $^{\Delta}$ DD. Residues are labeled according to the construct numbers (Fig. 3A). (C) Unassigned ^1H , ^{15}N -HSQC spectrum of ^{15}N VirFG $^{\Delta}$ DD alone. (D, E and F) Dynamical studies of VirFG $^{\Delta}$ DD. (D) Reference ^1H , ^{15}N heteronuclear NOE spectrum acquired at 600 MHz. (E) Saturated ^1H , ^{15}N heteronuclear NOE spectrum acquired at 600 MHz. (F) Resulting ^1H , ^{15}N heteronuclear NOE values plotted against unassigned peaks. (G) Unassigned ^1H , ^{15}N -HSQC spectrum of VirFG $^{\Delta}$ DD in complex with VirA $^{\Delta}$ DD. (H) Unassigned ^1H , ^{15}N -HSQC spectrum of VirFG $^{\Delta}$ DD in complex with VirA ACP₃₅- $^{\Delta}$ DD. (I) The final ensemble of 20 conformers that represents the solution structure of VirA $^{\Delta}$ DD-VirFG $^{\Delta}$ DD. VirA $^{\Delta}$ DD is colored red and VirFG $^{\Delta}$ DD blue. The introduced linker, N et C-termini are indicated. (J) Stereo-view of a representative VirA $^{\Delta}$ DD-VirFG $^{\Delta}$ DD structure depicted as cartoon. VirA $^{\Delta}$ DD is colored red and VirFG $^{\Delta}$ DD blue. The introduced linker, N et C-termini are indicated. (K) Dynamical study of VirA $^{\Delta}$ DD-VirFG $^{\Delta}$ DD construct. ^1H , ^{15}N heteronuclear NOE values are plotted against the sequence. The construct regions are colored as follows: red for VirA $^{\Delta}$ DD, gray for the linker and blue for VirFG $^{\Delta}$ DD. The N and C-termini are highly flexible as evidenced by negative ^1H , ^{15}N heteronuclear NOE. The linker region for which ^1H , ^{15}N heteronuclear NOEs are lower than 0.5 is also very dynamic. * = not possible to attribute the amide signal. (L) Superposition of the ^1H , ^{15}N -HSQC spectra of VirA $^{\Delta}$ DD in complex with VirFG $^{\Delta}$ DD (red), VirFG $^{\Delta}$ DD in complex with VirA $^{\Delta}$ DD (blue), and assigned VirA $^{\Delta}$ DD-VirFG $^{\Delta}$ DD (black).

S29

S30

Table S5. Summary of SAXS and SEC-MALLS data.

Construct	R_g (Guinier) (\AA) ^a	R_g (GNOM) (\AA) ^b	D_{max} (GNOM) (\AA) ^b	χ^2 (DAMMIN) ^c	χ^2 (CRYSOLE) ^c	Theoretical MW (kDa)	SAXS MoW (kDa) ^d (z range used)	MW by SEC-MALLS (kDa)
VirA $^{\Delta}$ DD	15.8 ± 0.01	16.38 ± 0.02	60	1.732–2.543	n.d.	4.30	8 (0.3)	8.2 ± 0.6
VirFG $^{\Delta}$ DD	16 ± 1	16.88 ± 0.01	66	2.321–2.489	6.61–10.39	4.11	5 (0.3)	4.2 ± 0.4
VirFG $^{\Delta}$ DD-linker	n.d.	n.d.	n.d.	n.d.	n.d.	6.87	n.d.	6.7 ± 0.6
VirA $^{\Delta}$ DD/VirFG $^{\Delta}$ DD complex	15.35 ± 0.01	15.76 ± 0.01	56	5.88	2.93	8.41 (1:1 complex)	8 (0.3)	n.d.
VirA $^{\Delta}$ DD-VirFG $^{\Delta}$ DD fusion	15.12 ± 0.01	15.49 ± 0.01	60	6.18	2.61	8.44	9 (0.4)	8.9 ± 0.9
VirA ACP ₃₅ - $^{\Delta}$ DD/VirFG $^{\Delta}$ DD complex	28.3 ± 0.3	31.0 ± 0.2	120	1.29	n.d.	18.2 (1:1 complex)	23 (0.3)	n.d.
VirA ACP ₃₅ - $^{\Delta}$ DD ³	37.0 ± 0.3	37.7	140	1.231	n.d.	14.3	31.4 ³	Rapid monomer/dimer exchange
VirA ACP ₃₅ - $^{\Delta}$ DD- $^{\Delta}$ DD VirFG fusion	27 ± 2	29.49 ± 0.01	118	2.44	n.d.	18.4	22 (0.25)	16.9 ± 0.9
VirA ACP ₃₅ - $^{\Delta}$ DD- $^{\Delta}$ DD-K ₅₅ VirFG fusion	48 ± 2	50.5 ± 0.1	200	1.76	n.d.	171.6 (dimer)	205 (0.3)	n.d.

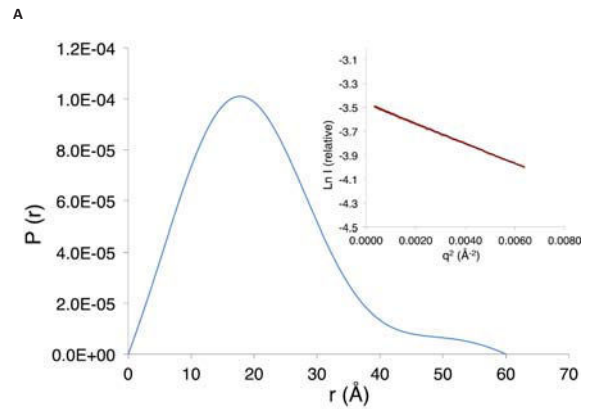
^a R_g is the radius of gyration derived from the atomic models using the program PRIMUS.⁵¹ For experimental data, R_g is given by the Guinier approximation.

^b D_{max} is the maximum particle diameter derived from the distance distribution function $P(r)$ using the program GNOM.⁵²

^c χ^2 is the discrepancy between the experimental SAXS profile and the theoretical curve calculated using the programs DAMMIN⁵³ or CRYSOLE.⁵⁴

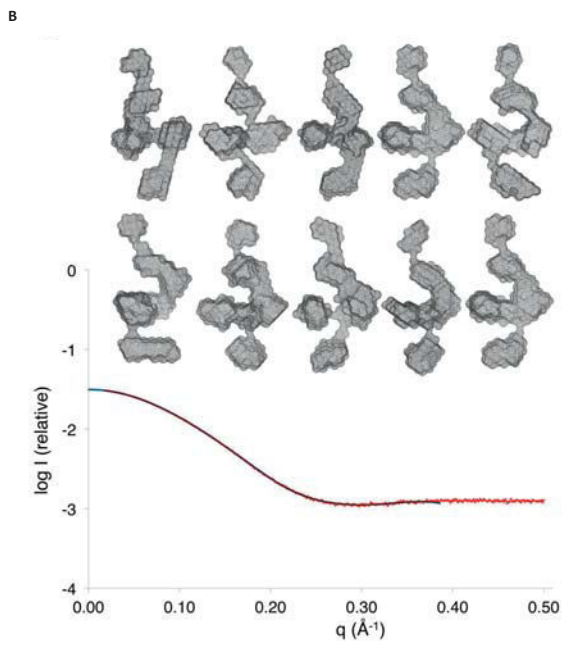
n.d.: not determined.

^dThe theoretical error on a typical molecular weight measurement by SAXS MoW is 10%.⁵⁵

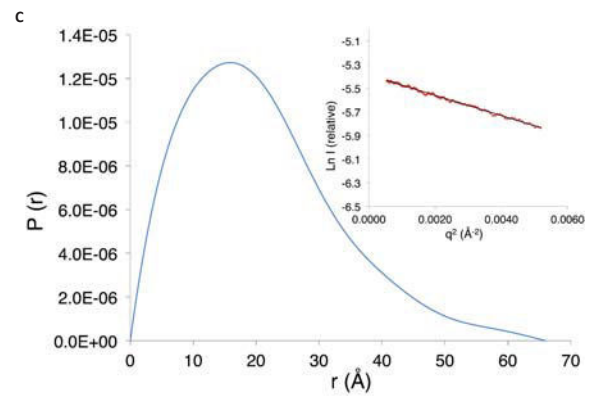


S31

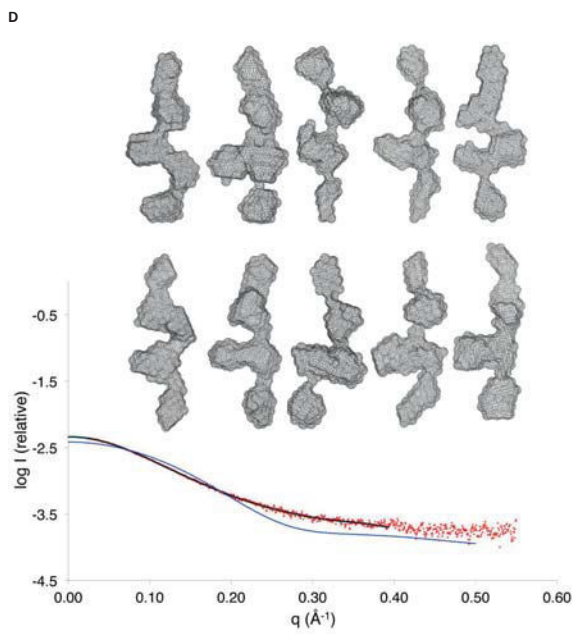
S32



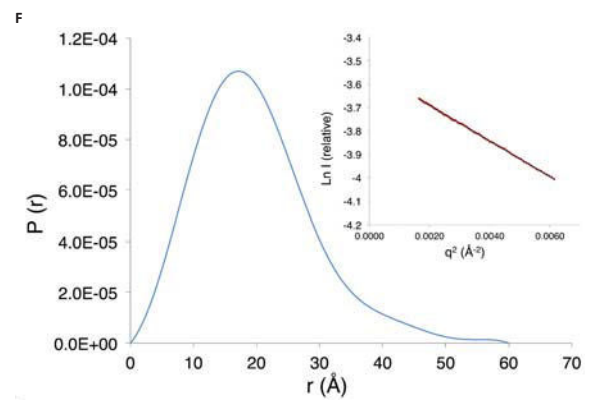
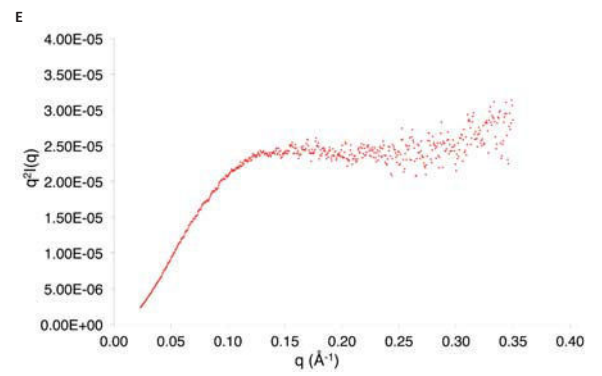
S33



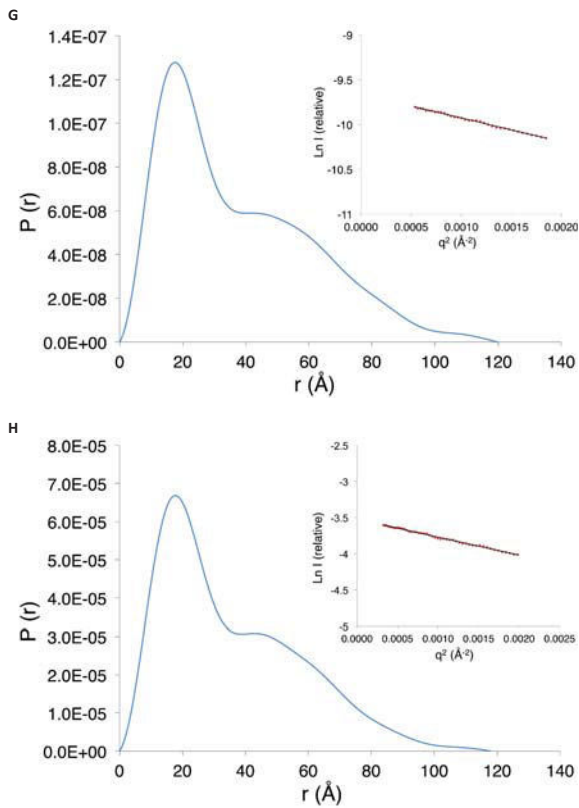
S34



S35



S36



S37

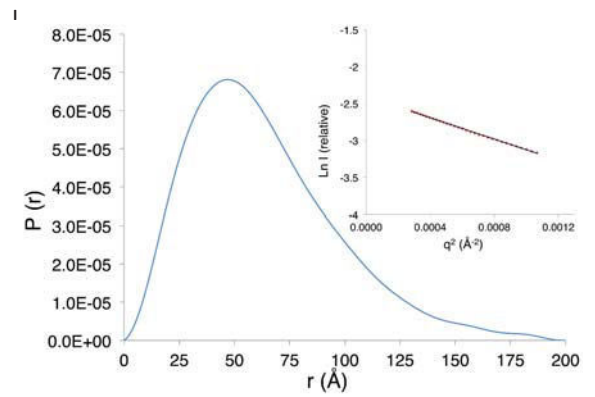
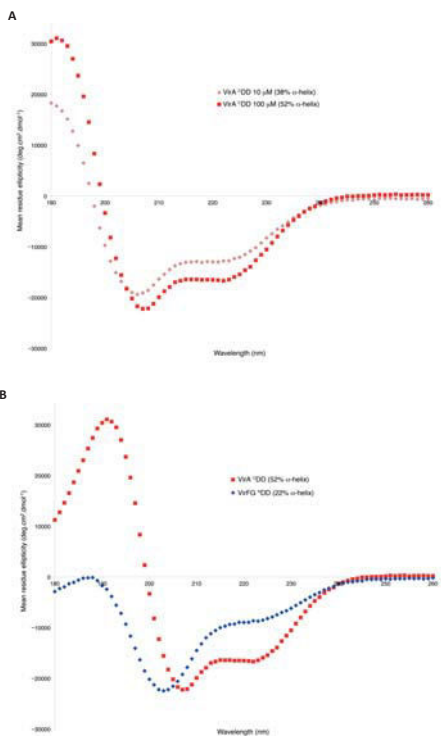
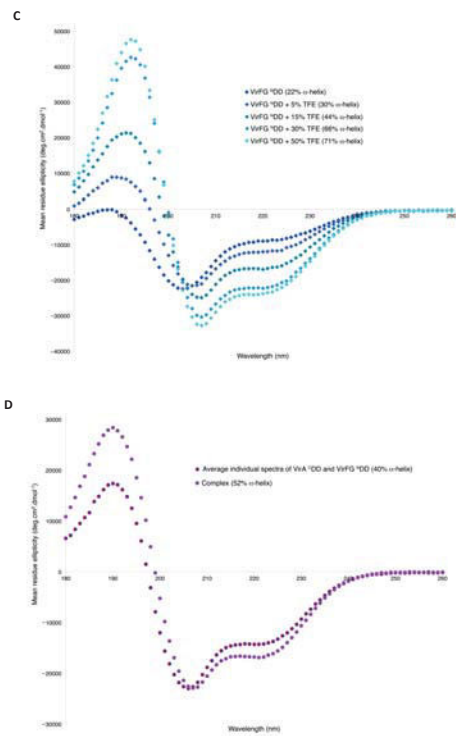


Figure S7. Analysis by SAXS of the various constructs. (A) The distance distribution function for VirA⁵DD calculated with GNOM. The Guinier plot is inset. (B) Panel of 10 representative *ab initio* form calculations for VirA⁵DD calculated with DAMMIN. The fit between the SAXS data and the 10 form calculations is shown ($\chi^2 = 1.732-2.543$). (C) The distance distribution function for VirFG⁵DD calculated with GNOM. The Guinier plot is inset. (D) Panel of 10 representative *ab initio* envelopes of VirFG⁵DD computed with DAMMIN. Superimposition of the experimental SAXS data acquired on VirFG⁵DD (red dots) and the theoretical curve for VirFG⁵DD calculated using CRYSOLO from one of the VirA⁵DD-VirFG⁵DD NMR structures among the 20 in the ensemble (solid blue line), shows that the two diverge significantly. In contrast, the fit between the 10 *ab initio* models (solid black line) and the experimental SAXS data (red dots) obtained on VirFG⁵DD is excellent ($\chi^2 = 2.321-2.489$). (E) The Kratky plot (red dots) for VirFG⁵DD calculated in the q range of 0.0005-0.35 Å⁻¹. (F) The distance distribution function for VirA⁵DD-VirFG⁵DD (covalent fusion) calculated with GNOM. The Guinier plot is inset. (G) The distance distribution function for VirA ACP₃₅₋₂DD/VirFG⁵DD (non-covalent complex) calculated with GNOM. The Guinier plot is inset. (H) The distance distribution function for VirA ACP₃₅₋₂DD-VirFG⁵DD (covalent complex) calculated with GNOM. The Guinier plot is inset. (I) The distance distribution function for VirA ACP₃₅₋₂DD-VirFG⁵DD-KS₆ (covalent complex) calculated with GNOM. The Guinier plot is inset.

S38



S39



S40

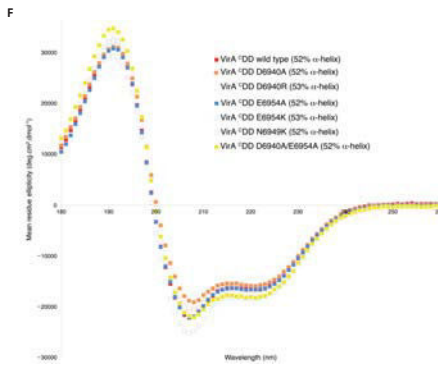
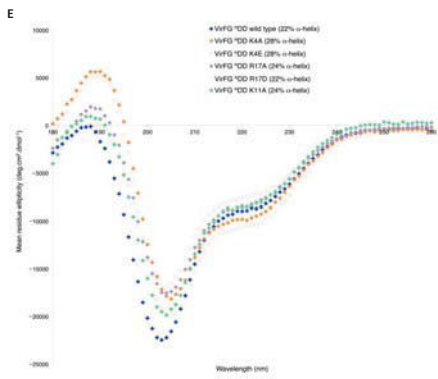


Figure S8. Analysis by circular dichroism of wild type and mutant docking domains. (A) Analysis of VirA^{CDD} at concentrations above (100 μ M) and below (10 μ M) the point at which it homodimerizes completely. The percentage α -helix is indicated in each case. (B) Analysis of wild type VirFG^{NDD} and for comparison, VirA^{CDD} (both at 100 μ M). The percentage α -helix is indicated in each case. (C) Effect of various concentrations of 2,2,2-trifluoroethanol (TFE) on the structure of VirFG^{NDD}. The percentage TFE and α -helix is indicated in each case. (D) Spectra showing net structuration upon combining VirA^{CDD}

and VirFG^{NDD}. A combined spectrum was obtained by analyzing simultaneously isolated VirA^{CDD} and VirFG^{NDD} (at 100 μ M). For comparison, the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μ M. The percentage α -helix is indicated in each case. (E) Analysis of site-directed mutants of VirFG^{NDD}. The introduced mutations resulted in substantial changes in secondary structure, as indicated by the calculated percentage α -helix. (F) Analysis of site-directed mutants of VirA^{CDD}. By and large, these mutations introduced minimal structural perturbation into the docking domain.

S41

S42

Table S6. NMR and refinement statistics for VirA^{CDD}-VirFG^{NDD}.

	Protein
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1372
Intra-residue ($ i-j =0$)	338
Inter-residue ($ i-j \neq 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
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.105 \pm 0.008
Max. distance constraint violation (Å)	0.12
Dihedral angle constraints (°)	0 \pm 0
Max. dihedral angle violation (°)	0
Deviations from idealized geometry	
Bond lengths (Å)	0.011 \pm 0.001
Bond angles (°)	1.947 \pm 0.024
Impropers (°)	n.d.
Average pairwise r.m.s. deviation* (Å)	
Heavy	VirA ^{CDD} 6936–6960, VirFG ^{NDD} 3–25
Backbone	0.99 \pm 0.13 0.38 \pm 0.09

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

Table S7. Putative interface residues in the VirA^{CDD}/VirFG^{NDD} docking domain families.

Interface	C-terminal docking domain →										N-terminal docking domain →													
	6939	6940	6942	6943	6945	6946	6949	6951	6953	6954	6956	6957	6958	6960	3	4	5	8	13	17	18	20	23	24
VirA ^{CDD}	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
VirFG ^{NDD}	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^C /C	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^F /F	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^D /D	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^S /S	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^R /R	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^E /E	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^G /G	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^H /H	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^I /I	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^J /J	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^K /K	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^L /L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^M /M	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^N /N	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^O /O	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^P /P	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^Q /Q	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^R /R	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^S /S	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^T /T	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^V /V	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^W /W	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^X /X	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^Y /Y	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Chia ^Z /Z	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L

*The first α -helix of the OzMH N-terminal docking domain is missing (confirmed by analysis of the published cluster sequence in the region of the gene (GenBank: AB590470.1)).

Docking domains present within the same assembly line are grouped and colored identically.

Key to residue types (as observed in the VirA/VirFG complex):

- Green/Orange = inter-docking domain hydrophobic interface
- Blue/Red = intra-docking domain hydrophobic interface
- Pink/Purple = inter-domain polar interactions

Residue 6956 participates in two types of interactions, as indicated.

S43

S44

SUPPLEMENTARY REFERENCES

- (1) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J.D.; Higgins, D.G. *Mol. Syst. Biol.* **2011**, *7*, 539.
- (2) Whicher, J.R.; Smaga, S.S.; Hansen, D.A.; Brown, W.C.; Gerwick, W.H.; Sherman, D.H.; Smith, J.L. *Chem. Biol.* **2013**, *20*, 1340.
- (3) Davison, J.; Dorival, J.; Rabeharindranto, H.; Mazon, H.; Chagot, B.; Gruez, A.; Weissman, K.J. *Chem. Sci.* **2014**, *5*, 3081.
- (4) Chang, Z.; Sitachitta, N.; Rossi, J.V.; Roberts, M.A.; Flatt, P.M.; Jia, J.; Sherman, D.H.; Gerwick, W.H. *J. Nat. Prod.* **2004**, *67*, 1356.
- (5) Partida-Martinez, L.P.; Hertweck, C. *ChemBioChem* **2007**, *8*, 41.
- (6) El-Sayed, A.K.; Hothersall, J.; Cooper, S.M.; Stephens, E.; Simpson, T.J.; Thomas, C.M. *Chem. Biol.* **2003**, *10*, 419.
- (7) Edwards, D.J.; Marquez, B.L.; Nogle, L.M.; McPhail, K.; Goeger, D.E.; Roberts, M.A.; Gerwick, W.H. *Chem. Biol.* **2004**, *11*, 817.
- (8) Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. *J. Biol. Chem.* **2002**, *277*, 13082.
- (9) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008**, *26*, 225.
- (10) Simunovic, V.; Zapp, J.; Rachid, S.; Meiser, P.; Müller, R. *ChemBioChem* **2006**, *7*, 1206.
- (11) Buntin, K.; Rachid, S.; Scharfe, M.; Blöcker, H.; Weissman, K.J.; Müller, R. *Angew. Chem. Int. Ed Engl.* **2008**, *47*, 4595.
- (12) Dehn, R.; Katsuyama, Y.; Weber, A.; Gerth, K.; Jansen, R.; Steinmetz, H.; Höfle, G.; Müller, R.; Kirschning, A. *Angew. Chem. Int. Ed Engl.* **2011**, *50*, 3882.
- (13) Pulsawat, N.; Kitani, S.; Nihira, T. *Gene* **2007**, *393*, 31.
- (14) Mast, Y.; Weber, T.; Götz, M.; Ort-Winklbauer, R.; Gondran, A.; Wohlleben, W.; Schinko, E. *Microb. Biotechnol.* **2011**, *4*, 192.
- (15) Alt, S.; Wilkinson, B. *ACS Chem. Biol.* **2015**, *10*, 2468.
- (16) Jungmann, K.; Jansen, R.; Gerth, K.; Huch, V.; Fenical, W.; Müller, R. *ACS Chem. Biol.* **2015**, *10*, 2480.
- (17) Perlova, O.; Gerth, K.; Kaiser, O.; Hans, A.; Müller, R. *J. Biotechnol.* **2006**, *121*, 174.
- (18) Erol, O.; Schäberle, T.F.; Schmitz, A.; Rachid, S.; Gurgui, C.; El Omari, M.; Lohr, F.; Kehraus, S.; Piel, J.; Müller, R.; König, G.M. *ChemBioChem* **2010**, *11*, 1253.
- (19) Zhang, F.; He, H.Y.; Tang, M.C.; Tang, Y.M.; Zhou, Q.; Tang, G.L. *J. Am. Chem. Soc.* **2011**, *133*, 2452.
- (20) Weber, T.; Laiple, K.J.; Pross, E.K.; Textor, A.; Grond, S.; Welzel, K.; Pelzer, S.; Vente, A.; Wohlleben, W. *Chem. Biol.* **2008**, *15*, 175.
- (21) Matilla, M.A.; Stöckmann, H.; Leeper, F.J.; Salmond, G.P. *J. Biol. Chem.* **2012**, *287*, 39125.
- (22) Zhao, C.; Coughlin, J.M.; Ju, J.; Zhu, D.; Wendt-Pienkowski, E.; Zhou, X.; Wang, Z.; Shen, B.; Deng, Z. *J. Biol. Chem.* **2010**, *285*, 20097.
- (23) Wang, C.; Henkes, L.M.; Doughty, L.B.; He, M.; Wang, D.; Meyer-Almes, F.J.; Cheng, Y.Q. *J. Nat. Prod.* **2011**, *74*, 2031.
- (24) Fukuda, D.; Haines, A.S.; Song, Z.; Murphy, A.C.; Hothersall, J.; Stephens, E.R.; Gurney, R.; Cox, R.J.; Crosby, J.; Willis, C.L.; Simpson, T.J.; Thomas, C.M. *PLoS One* **2011**, *6*, e18031.
- (25) Mahenthiralingam, E.; Song, L.; Sass, A.; White, J.; Wilmot, C.; Marchbank, A.; Boisha, O.; Paine, J.; Knight, D.; Challis, G.L. *Chem. Biol.* **2011**, *18*, 665.
- (26) Wakimoto, T.; Egami, Y.; Nakashima, Y.; Wakimoto, Y.; Mori, T.; Awakawa, T.; Ito, T.; Kenmoku, H.; Asakawa, Y.; Piel, J.; Abe, I. *Nat. Chem. Biol.* **2014**, *10*, 648.
- (27) Kampa, A.; Gagunashvili, A.N.; Gulder, T.A.; Morinaka, B.I.; Daolio, C.; Godejohann, M.; Miao, V.P.; Piel, J.; Andrésson, Ö. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E3129.
- (28) Piel, J.; Hui, D.; Wen, G.; Butzke, D.; Platzer, M.; Fusetani, M.; Matsunaga, S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16222.
- (29) Piel, J. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14002.
- (30) Fisch, K.M.; Gurgui, C.; Heycke, N.; van der Sar, S.A.; Anderson, S.A.; Webb, V.L.; Taudien, S.; Platzer, M.; Rubio, B.K.; Robinson, S.J.; Crews, P.; Piel, J. *Nat. Chem. Biol.* **2009**, *5*, 494.
- (31) Sudek, S.; Lopanik, N.B.; Waggoner, L.E.; Hildebrand, M.; Anderson, C.; Liu, H.; Patel, A.; Sherman, D.H.; Haygood, M.G. *J. Nat. Prod.* **2007**, *70*, 67.
- (32) Menche, D.; Arikian, F.; Perlova, O.; Horstmann, N.; Ahlbrecht, W.; Wenzel, S.C.; Jansen, R.; Irschik, H.; Müller, R. *J. Am. Chem. Soc.* **2008**, *130*, 14234.
- (33) Butcher, R.A.; Schroeder, F.C.; Fischbach, M.A.; Straight, P.D.; Kolter, R.; Walsh, C.T.; Clardy, J. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1506.
- (34) Mattheus, W.; Gao, L.J.; Herdewijn, P.; Landuyt, B.; Verhaegen, J.; Masschelein, J.; Volckaert, G.; Lavigne, R. *Chem. Biol.* **2010**, *17*, 149.
- (35) Carvalho, R.; Reid, R.; Viswanathan, N.; Gramajo, H.; Julien, B. *Gene* **2005**, *359*, 91.
- (36) Kopp, M.; Irschik, H.; Pradella, S.; Müller, R. *ChemBioChem* **2005**, *6*, 1277.
- (37) Pistorius, D.; Müller, R. *ChemBioChem* **2012**, *13*, 416.
- (38) Chen, X.-H.; Vater, J.; Piel, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsis, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A.W.; Gottschalk, G.; Süßmuth, R.D.; Borriss, R. *J. Bacteriol.* **2006**, *188*, 4024.
- (39) Cheng, Y.-Q.; Tang, G.-L.; Shen, B. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3149.
- (40) Ueoka, R.; Uria, R.; Reiter, S.; Mori, T.; Karbaum, P.; Peters, E.E.; Helfrich, E.J.; Morinaka, B.I.; Gugger, M.; Takeyama, H.; Matsunaga, S.; Piel, J. *Nat. Chem. Biol.* **2015**, *11*, 705.
- (41) Schneider, K.; Chen, X. H.; Vater, J.; Franke, P.; Nicholson, G.; Borriss, R.; Süßmuth, R.D. *J. Nat. Prod.* **2007**, *70*, 1417.
- (42) Irschik, H.; Kopp, M.; Weissman, K.J.; Buntin, K.; Piel, J.; Müller, R. *ChemBioChem* **2010**, *11*, 1840.
- (43) Liu, X.; Biswas, S.; Berg, M.G.; Antapli, C.M.; Xie, F.; Wang, Q.; Tang, M.C.; Tang, G.L.; Zhang, L.; Dreyfuss, G.; Cheng, Y.Q. *J. Nat. Prod.* **2013**, *76*, 685.
- (44) Moebius, N.; Ross, C.; Scherlach, K.; Rohm, B.; Roth, M.; Hertweck, C. *Chem. Biol.* **2012**, *19*, 1164.
- (45) Seo, J.-W.; Ma, M.; Kwong, T.; Ju, J.; Lim, S.K.; Jiang, H.; Lohman, J.R.; Yang, C.; Cleveland, J.; Zazopoulos, E.; Farnet, C.M.; Shen, B. *Biochemistry* **2014**, *53*, 7854.
- (46) Arakawa, K.; Sugino, F.; Kodama, K.; Ishii, T.; Kinashi, H. *Chem. Biol.* **2005**, *12*, 249.
- (47) Lim, S.-K.; Ju, J.; Zazopoulos, E.; Jiang, H.; Seo, J.W.; Chen, Y.; Feng, Z.; Rajsiki, S.R.; Farnet, C.M.; Shen, B. *J. Biol. Chem.* **2009**, *284*, 29746.
- (48) Royer, M.; Costet, L.; Vivien, E.; Bes, M.; Cousin, A.; Damais, A.; Piretti, I.; Savin, A.; Megessier, S.; Viard, M.; Frutos, R.; Gabriel, D.W.; Rott, P.C. *Mol. Plant Microbe Interact.* **2004**, *17*, 414.
- (49) Ross, A.C.; Xu, Y.; Lu, L.; Kersten, R.D.; Shao, Z.; Al-Suwalleh, A.M.; Dorrestein, P.C.; Qian, P.Y.; Moore, B.S. *J. Am. Chem. Soc.* **2013**, *135*, 1155.
- (50) Pei, J.; Grishin, N.V. *Methods Mol. Biol.* **2014**, *1079*, 263.
- (51) Konarev, P.V.; Volkov, V.V.; Sokolova, A.V.; Koch, M.H.J.; Svergun, D.I. *J. Appl. Crystallogr.* **2003**, *36*, 1277.
- (52) Svergun, D. *J. Appl. Crystallogr.* **1992**, *25*, 495.
- (53) Svergun, D.I. *Biophys. J.* **1999**, *76*, 2879.
- (54) Svergun, D.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, *28*, 768.
- (55) Fischer, H.; de Oliveira Neto, M.; Napolitano, H.B.; Polikarpov, I.; Craievich, A.F. *J. Appl. Cryst.* **2010**, *43*, 101.

APPENDIX

Characterization of Intersubunit Communication in the Virginiamycin *trans*-Acyl Transferase Polyketide Synthase

Jonathan Dorival,[†] Thibault Annaval,[†] Fanny Risser,[†] Sabrina Collin,[†] Pierre Roblin,^{†,5} Christophe Jacob,[†] Arnaud Gruez,^{†,†} Benjamin Chagot,^{†,†} and Kira J. Weissman,^{†,7}

[†]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

[†]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 Gif-sur-Yvette CEDEX, France

⁵UR1268 Biopolymères, Interactions Assemblages (BIA), INRA, Rue de la Géraudière BP 71627, 44316 NANTES CEDEX 3, France

We aimed to strengthen the case that our observations concerning the VirA^ΔDD/VirFG^ΔDD interface are generalizable to other *trans*-AT PKSs. For this, we benchmarked several web-based servers for the *in silico* identification of IDPs and protein-protein interaction regions (IUPred/ANCHOR^{1,2}, PONDR^{3,4}, DisoPred⁵ and DisEMBL⁶) against VirA^ΔDD and VirFG^ΔDD, 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 proteins⁷) and so a combination of programs should give a more confident prediction. We also analyzed the sequences using a coiled-coil predictor⁸ (green = window of 14 residues; blue = window of 21 residues; red = window of 28 residues), which we validated against the known coiled-coil region of a Class 1 N-terminal DDs^{9,10} (page 29). (Note: coiled-coil regions are predicted to be disordered by PONDR, as the monomers only fold correctly in the presence of partner). 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^ΔDD interaction region (residues Q6936–L6960) is largely structured, while ANCHOR identified the two α -helices as binding motifs (data provided on page 2). Furthermore, no coiled-coil was predicted for the sequence. In the case of VirFG^ΔDD (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 (page 2). Again, the coiled-coil prediction was negative. Gratifyingly, analogous results were obtained for the regions corresponding to VirA^ΔDD and VirFG^ΔDD from the homologous pristinamycin PKS (SnaE^ΔDD and SnaE3^ΔDD, respectively (page 3)), despite overall low levels of sequence identity between the docking regions (53% between VirA^ΔDD and SnaE2^ΔDD and 39% between VirFG^ΔDD and SnaE3^ΔDD¹¹).

We next extended this analysis to the remaining 25 putative VirA^ΔDD/VirFG^ΔDD-type docking interfaces from *trans*-AT PKSs assigned on the basis of sequence analysis and secondary structure prediction (data provided on pages 5–31, with the data for each matched pair of docking domains shown on a single page). For 10 of these pairs (indicated in pink in the table below), the predictions were in good to excellent agreement with those for VirA^ΔDD and VirFG^ΔDD (analysis of each pair relative to VirA^ΔDD and VirFG^ΔDD is presented on each and highlighted in pink), while for 3 others, only the N-terminal docking domain shares predicted properties with VirFG^ΔDD. Another important finding for 9 interfaces was the indication that the N-terminal partner forms a coiled-coil; this implies the existence of another structural class of docking domains within *trans*-AT PKS (in fact, the majority of these pairs show convincing homology to the Class 2 docking domains of *cis*-AT PKS (see Figure 1, page 31)).

Couples analyzed:

C-terminal docking domain	N-terminal docking domain	C-terminal docking domain	N-terminal docking domain
VirA	VirFG	KirAV	KirAV (coiled-coil?)
SnaE2	SnaE3	MlnD	MlnE
AtcE	AtcF	OzmN	OzmH
BaeM	BaeN (coiled-coil?)	OzmH	OzmJ
Bat 1	Bat2	OzmK	OzmL
BryA	BryB (coiled-coil?)	RhC	RhD (coiled-coil?)
BryB	BryC (coiled-coil?)	RizC	RizD (coiled-coil?)
ChiB	ChiC	Ta1	TaO (coiled-coil?)
ChiC	ChiD	TaiD	TaiE
ChiE	ChiF	TaiE	TaiK
DisB	DisC	TaiL	TaiM
ElaK	ElaO	5923	5922 (coiled-coil?)
EtnH	EtnI (coiled-coil?)	5925	5924
KirAll	KirAllI		

1

2

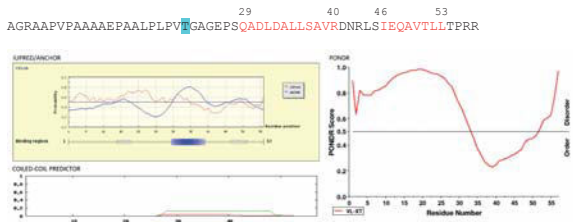
Key:

Red = For VirA^ΔDD/VirFG^ΔDD observed α -helices in the complex; for all others, α -helices predicted by PspPred

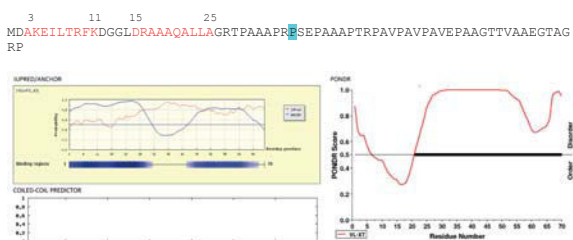
Blue = start/end site of expression constructs for VirA^ΔDD/VirFG^ΔDD, respectively

All sequences are numbered starting from 1 to match the output of the programs. Note, some N-terminal DDs were analyzed with the first M present, while others were not – in the latter cases, the missing M is indicated as (M).

VirA^ΔDD

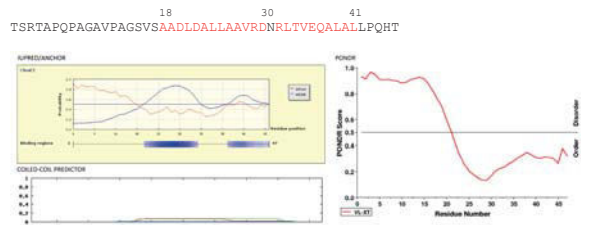


VirFG^ΔDD



3

SnaE2^ΔDD

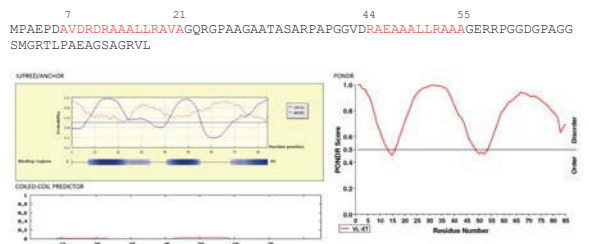


OBSERVATIONS

Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is not predicted either to be disordered (IUPred/PONDR) or to form a coiled-coil.

Strong agreement with VirA^ΔDD

SnaE3^ΔDD



OBSERVATIONS

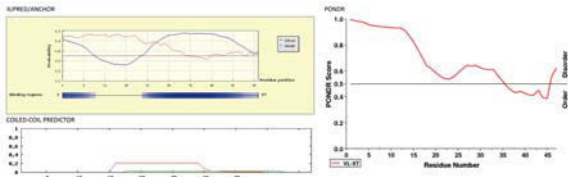
Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be largely disordered (IUPred/PONDR) but not to form a coiled-coil.

Strong agreement with VirFG^ΔDD

4

AtcE^ΔDD

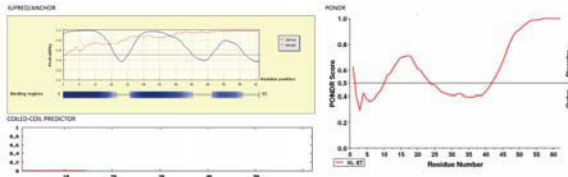
HAGHAAAAPPAGDAGGERATDVEDMLRRLAAGELTVAEAYNLGGPA



OBSERVATIONS
 Predicted interaction region (ANCHOR) overlaps well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred/PONDRL) and not to form a coiled-coil.
 Agreement with VirA^ΔDD

AtcF^ΔDD

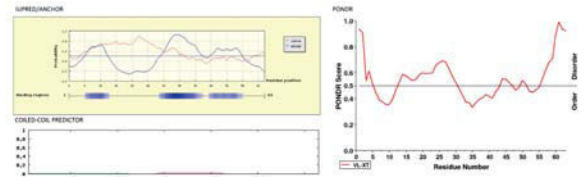
(M) NSAVKRVLELVETRQISPEEGRRRLTALGTAGDRPVAGEADRPRPFDGRAAAPASEGPPSD



OBSERVATIONS
 Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be largely disordered (IUPred/PONDRL), but not to form a coiled-coil.
 Agreement with VirFG^ΔDD

BaeM^ΔDD

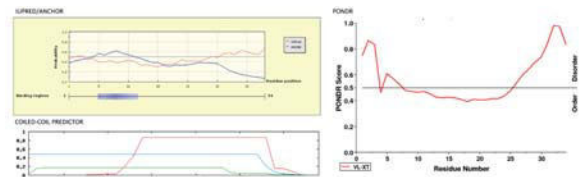
LTEHKETIETFYRTEETEETEAAPESKEYTDQEI IAMKQVSEGLDFKRQVDIIEGSKTYES



OBSERVATIONS
 Predicted interaction regions (ANCHOR) overlap well with the predicted α -helices. The docking region is predicted to be borderline disordered (IUPred/PONDRL), and not to form a coiled-coil.
 Agreement with VirA^ΔDD

BaeN^ΔDD

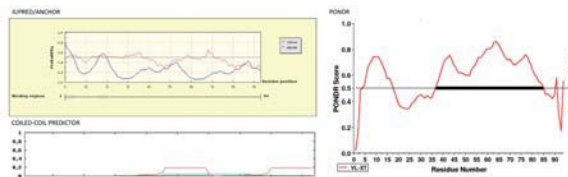
(M) NHKELLDAYRSGTLTIAEVEQKQLQAFKRTTAKRP



OBSERVATIONS
 The first predicted α -helix is predicted to be an interaction region (ANCHOR). The docking region is predicted to be borderline disordered (IUPred/PONDRL), but to form a coiled-coil.
 Does not agree with VirFG^ΔDD; may be an alternative type of docking domain?

Bat1^ΔDD

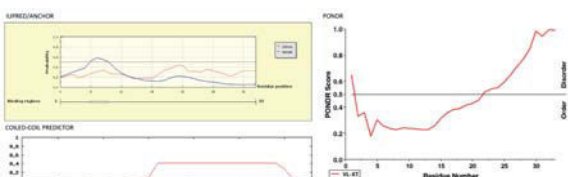
VRIVISQPREASSARTLDVVVGERLDTQSSDDRTMKMLEDFRDGNVKEEVEGILDRWTKQSSDDRT
 MKMLEDFFDGNVKEEVEGILDRWTK



OBSERVATIONS
 The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be disordered (IUPred/PONDRL), and not to form a coiled-coil.
 Poor agreement with VirA^ΔDD

Bat2^ΔDD

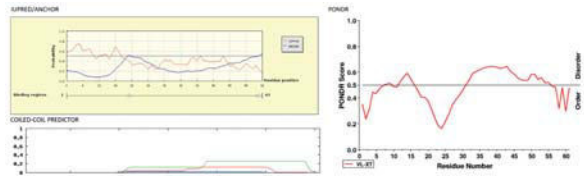
(M) TRRDLLFAYKNRSLTQELVAALTELNTERRFSS



OBSERVATIONS
 The secondary structure elements are not predicted to be interaction regions (ANCHOR). The docking region is not predicted to be disordered (IUPred/PONDRL), and not strongly predicted to form a coiled-coil.
 Poor agreement with VirFG^ΔDD

BryA^ΔDD

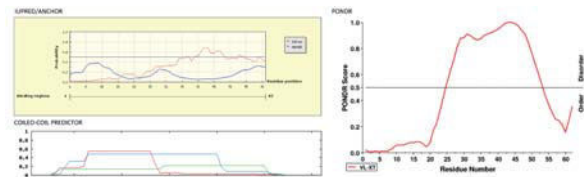
RKQIISNDEKQHQPSISTIFPPTLDELLKKIQEGTLGIEAADQLIDELPDYHLMDELHELL



OBSERVATIONS
 The predicted α -helices are not predicted to be an interaction region (ANCHOR). The docking region is predicted to be borderline (IUPred) to mixed ordered/disordered (PONDRL), and not to form a coiled-coil.
 Poor agreement with VirA^ΔDD

BryB^ΔDD

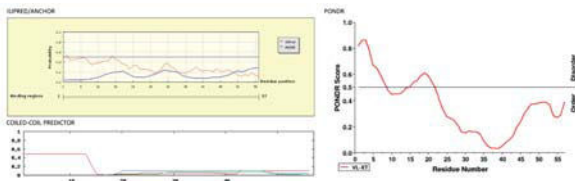
(M) VVVEEFFVSYRDILKALQDEKISFEEAKYKLIKRRDKKSKQLRNLHDELNRSMKITPKIVNN



OBSERVATIONS
 The predicted α -helices are not predicted to be interaction regions (ANCHOR). The first α -helix is predicted to be ordered and the second disordered (IUPred/PONDRL), and the overall domain to form a coiled-coil.
 Does not agree with VirFG^ΔDD; may be an alternative type of docking domain?

BryB^oDD

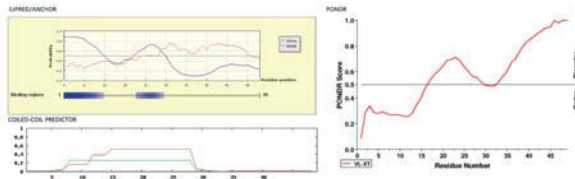
RQKIEQKNDIQNLSSSYFSEDNQ**EV**ALSLKEILKKVEGNTLS**ID**KA**E**LF**EQ**FLSK

**OBSERVATIONS**

The predicted α -helices are not predicted to be an interaction region (ANCHOR). The docking region is predicted to be ordered (IUPred/PONDR), and not to form a coiled-coil. Poor agreement with VirA^oDD

BryC^oDD

(M) KAKV**I**F**Q**GLKS**N**E**I**S**I****E**Q**A**E**K**M**L**L**S**L**D**T**E**S**K**H**G**H**F**E**S**K**L**S**D**I**K**S**L**D**L**H**D**D

**OBSERVATIONS**

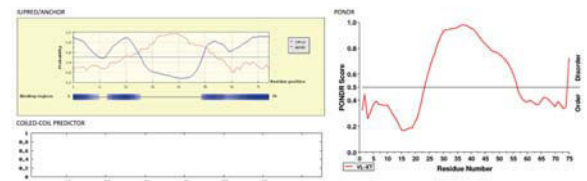
The predicted secondary structure elements are predicted to be interaction regions (ANCHOR). The docking region is predicted to be largely ordered (IUPred/PONDR), and there is some indication for coiled-coil formation.

Does not agree with VirFG^oDD; may be an alternative type of docking domain?

9

ChiB^oDD

LSSFAADVRLPGAAKIDTAALFGDGAAPAPAPAPVDFPRGGEPEGDE**FL**WRLLRGV**ES**GA**L**D**V**NEA
CRLLLELP
70

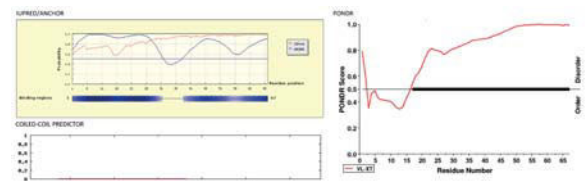
**OBSERVATIONS**

Predicted interaction region (ANCHOR) overlaps well with predicted α -helices. The docking region is not predicted to be disordered (IUPred/PONDR) nor to form a coiled-coil.

Strong agreement with VirA^oDD

ChiC^oDD

(M) KR**D**PR**V**IF**E**Q**L**KN**G**Q**I**S**R**AE**A**HR**L**L**K**ARS**D**GRE**Q**E**A**T**G**PGA**E**GAA**E**AR**A**PS**A**PP**Q**P**A**SS**D**S**D**
VG

**OBSERVATIONS**

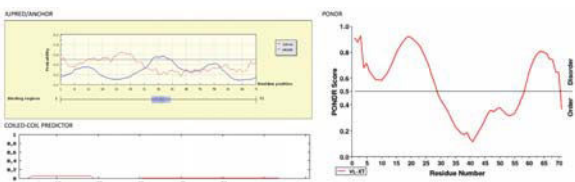
Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred/PONDR), but not to form a coiled-coil.

Strong agreement with VirFG^oDD

10

ChiC^oDD

ERDPALAEIKAVLRAQGA**P**GA**P**ELAV**G**H**D**EL**W**RL**L**GA**L**GA**D**EL**D**AL**GA**EL**D**AL**F**N**L**R**S**R**P**E**L**E
GMV

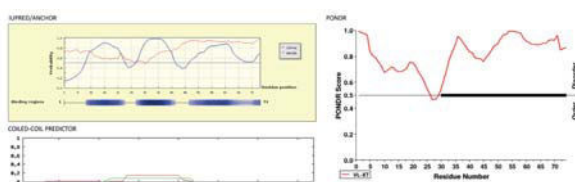
**OBSERVATIONS**

Predicted α -helices not confidently predicted as interaction regions. The C-terminal extreme of the docking region is predicted to be disordered (IUPred/PONDR), while the whole region is not predicted to form a coiled-coil.

Relatively poor agreement with VirA^oDD

ChiD^oDD

MQR**D**RG**R**RL**Y**ME**K**AV**T**D**A**M**F**D**E**RS**E**E**K**LS**V**F**R**AL**R**LR**R**TP**P**Q**H**GA**D**GA**P**RA**V**L**P**AD**P**Q**N**
RHR**P**FP

**OBSERVATIONS**

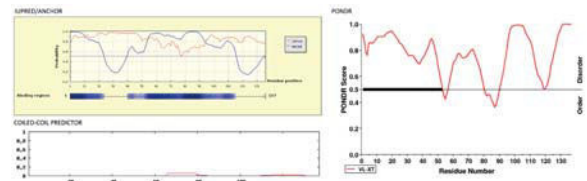
Predicted interaction regions (ANCHOR) overlap well with predicted α -helix. The docking region is strongly predicted to be disordered (IUPred/PONDR), but not to form a coiled-coil.

Strong agreement with VirFG^oDD

11

ChiE^oDD

V**D**H**E**A**L**LR**A**SE**P**PG**V**GG**G**AE**P**V**R**FR**A**NP**R**PER**F**ER**A**D**P**PER**A**GV**A**PG**R**RA**P**SP**P**AA**A**VER**F**D**G**AV**P****E**
L**A**GE**V**AR**M**S**D**A**E**VER**L**L**G**W**L**Q**G**G**A**G**A**G**A**G**A**G**A**T**N**SS**A**VR**A**S**L**D**D****Q**Q**E**RE**A**RR**R**L**D**R**G**P**S**GE**W**
R

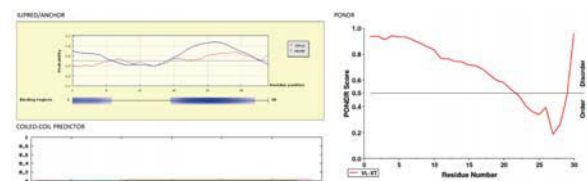
**OBSERVATIONS**

Predicted α -helices predicted as interaction regions. The docking region is predicted to be disordered (IUPred/PONDR), and not predicted to form a coiled-coil.

Some agreement with VirA^oDD

ChiF^oDD

MAM**T**L**R**S**Q**DL**M**DA**K**RES**L**V**R**LL**H**Q**K**AN**R**

**OBSERVATIONS**

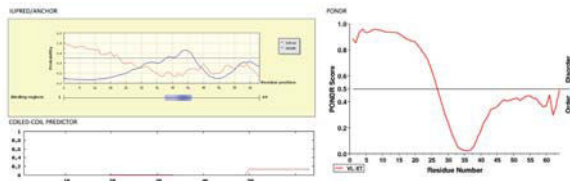
A portion of the predicted interaction regions (ANCHOR) overlaps well with the predicted α -helix. The docking region is predicted to be disordered (IUPred/PONDR), but not to form a coiled-coil.

Agreement with VirFG^oDD

12

DisB^oDD

20 46 51 62
 SRRGARAPESRAKARSGAAGADLSTSLKALSGAVLREQFLAFGHDLAGVPGEEELTRLYAILQEE

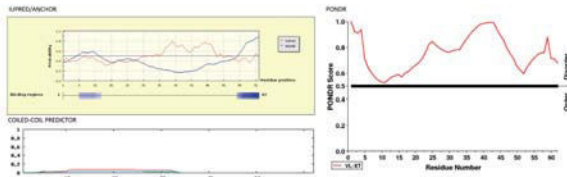


OBSERVATIONS

The first predicted α -helix is predicted to be an interaction region (ANCHOR), though not the second. The docking region is predicted to be largely ordered (IUPred/PONDR), and not to form a coiled-coil. [Agreement with VirA^oDD](#)

DisC^oDD

6 11 15 30
 MESAMTIQEFANLSAEKVQVLRDRRASWQAPEGPAASAQPSLRPFVITARPGRFLFPF

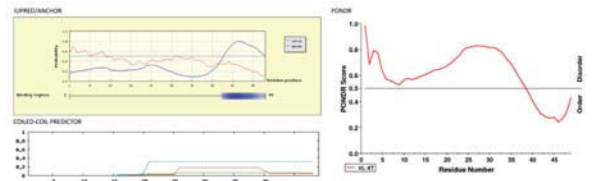


OBSERVATIONS

The first predicted α -helix is predicted to be an interaction region (ANCHOR), though not the second. The docking region is predicted to be borderline (IUPred) to strongly disordered (PONDR), and not predicted to form a coiled-coil. [Agreement with VirFG^oDD](#)

ElaK^oDD

22 31 33 47
 SSEMTNISTYEPVQSLQQEFLDDVVVRQVEKEMDAEAAANLLAVLSKQ

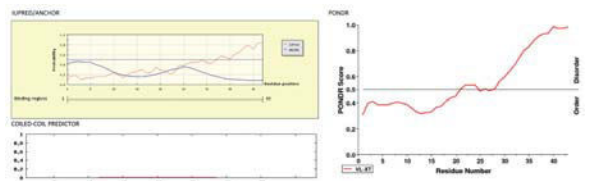


OBSERVATIONS

Second α -helix predicted as interaction region. The first α -helix is predicted to be borderline to disordered while the second is ordered (IUPred/PONDR), and neither is strongly predicted to form a coiled-coil. Some agreement with VirA^oDD

ElaO^oDD

2 10 17 26
 (M) DSLAILKALSTGAISAGAAREALKQLAVAPVSVKPGHDARATP

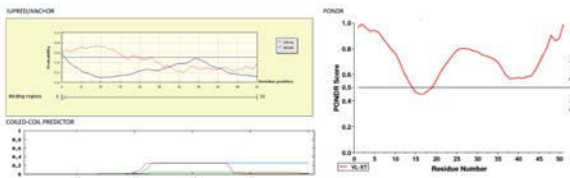


OBSERVATIONS

The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be borderline disordered (IUPred/PONDR), but not to form a coiled-coil. Some agreement with VirFG^oDD

EtnH^oDD

20 49
 CAEHPPDLARRQAGTQANGGAVLERRAVEALAEELAGRISLDQTLRLRA

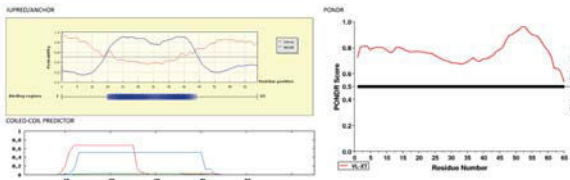


OBSERVATIONS

The predicted α -helix is not predicted to be an interaction region (ANCHOR). The docking region is predicted to be ordered (IUPred) to disordered (PONDR), and not to form a coiled-coil. Poor agreement with VirA^oDD

EtnI^oDD

10 43
 (M) FEHSYEPFRGESTVKSQIEALLREYQSGLLSEAE LRARYQLVAGSDGAPARPADRAGATQTVRR

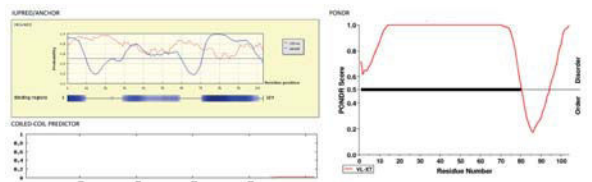


OBSERVATIONS

The predicted α -helix is predicted to be an interaction region (ANCHOR). The docking region is predicted to be borderline (IUPred) to strongly disordered (PONDR), and there is some indication for coiled-coil formation. Does not agree with VirFG^oDD; may be an alternative type of docking domain?

KirAll^oDD

72 85 91 99
 HDKYGLPATATPEPPPAPVTSSATAPEPTSVFVASSAAAPAPVAPVAPSAAPAPVASSAAPEPP
 PAPVADAVGALLAGLRDGSVTDDALAE LRQGNVR

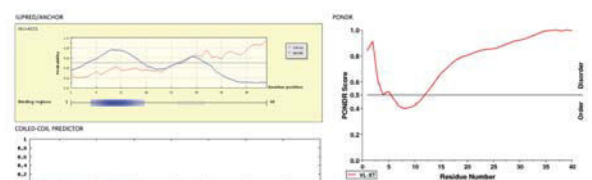


OBSERVATIONS

Predicted interaction region (ANCHOR) overlaps well with predicted α -helices. The docking region is predicted to be mixed ordered/disordered (IUPred/PONDR), and not to form a coiled-coil. [Agreement with VirA^oDD](#)

KirAllI^oDD

3 13 18 29
 MSRALRDVLAAVAAGELPDEIAESLLRGLSDPVPEPGYPF

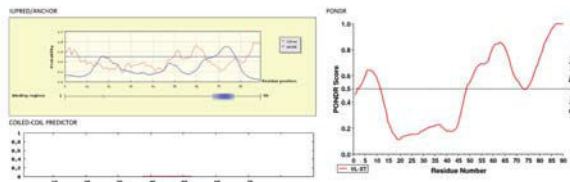


OBSERVATIONS

Predicted interaction regions (ANCHOR) overlap well with predicted α -helices (though the second interaction region isn't strong). The docking region is predicted to be borderline disordered (IUPred/PONDR), but not to form a coiled-coil. [Agreement with VirFG^oDD](#)

KirAV^ΔDD

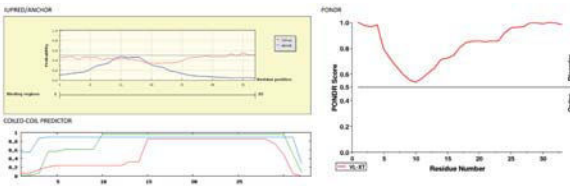
AGRFDGGTGGADLPLAFADFRLAAGLLEAAATDEVFRDRVVERVRLTALSATPEAP**PDLFAAAT**
 DDEV**FALIDAE**LGSFVRGQEDR
 71 79



OBSERVATIONS
 The second α -helix is predicted to be an interaction region (ANCHOR). The docking region is predicted to be mixed ordered/disordered (IUPred/PONDR), and not to form a coiled-coil. Some agreement with VirA^ΔDD

KirAVI^ΔDD

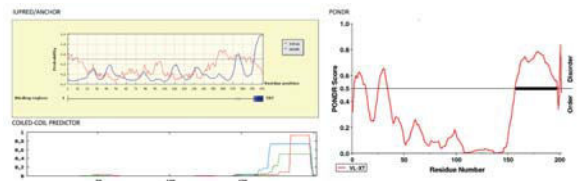
M**SS**EDKLRDYLRRVTAELQVTRRRRLREVESRSG
 4 31



OBSERVATIONS
 The predicted α -helix is not predicted to be an interaction region. The docking region is predicted to be disordered (IUPred/PONDR) and strongly predicted to form a coiled-coil. Does not agree with VirFG^ΔDD; may be an alternative type of docking domain?

MinD^ΔDD

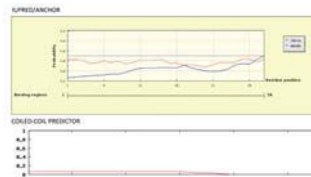
SSIRNRSADKPYNGNFRKSGFGLLEEIAYTLQGTREFFECRLAITSNIGD**LDLEK**LNGYIQGSR**LQGV**
 YEGNVTFDFSLYAPLFGGPEGHAYVQSLIQ**QNKPAKLAQL**WTIGCEISWEDLYEELKPKV**KL**SLPLY
 FSKNRHWIHLNKDEAKAEAFSNTELHNSNKDTADGEIMSLLEKAQTGEMMKETS**ELIE**LLFHE
 96 110



OBSERVATIONS
 The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be strongly ordered (IUPred/PONDR), and not to form a coiled-coil. Poor agreement with VirA^ΔDD

MinE^ΔDD

(M)NE**KTMQ**IEQIVK**KVKE**CSLT**PEE**GLELI
 3 16 21 23

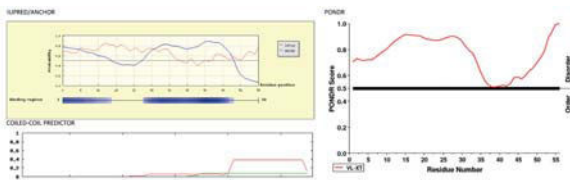


Sequence is too short for PONDR

OBSERVATIONS
 The predicted α -helices are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be borderline ordered (IUPred/PONDR), and not to form a coiled-coil. Poor agreement with VirA^ΔDD

OzmN^ΔDD

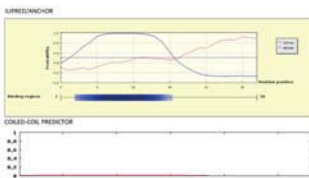
AANTGAAGTAAGTAAGPARAAE**FEAGIE**QLVDSLSDTEVDSLRLQ**LG**SVLQ**KQEE**QR
 22 54



OBSERVATIONS
 Predicted interaction region (ANCHOR) overlaps well with the predicted α -helix. The docking region is predicted to be borderline to disordered (IUPred/PONDR), and not to form a coiled-coil. Agreement with VirA^ΔDD

OzmH^ΔDD

MRRRS**MDDK**RALLKLLASVR**QQA**APATAPRRTA
 6 24

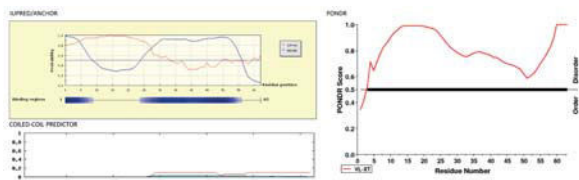


Too short for PONDR

OBSERVATIONS
 Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred), but not to form a coiled-coil. Strong agreement with VirFG^ΔDD

OzmH^ΔDD

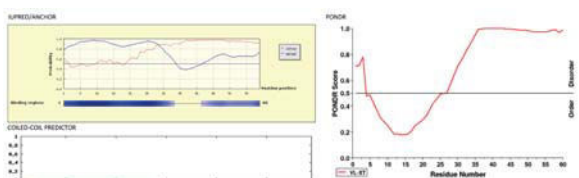
IELVGPDEPTAGSGEPPADGTGDTAA**PDELI**AKLEAR**FAAG**ELSA**AEVLD**LLD**LAEL**ATRERR
 28 43 46 60



OBSERVATIONS
 Predicted interaction region (ANCHOR) overlaps well with the predicted α -helices. The docking region is predicted to be borderline to disordered (IUPred/PONDR), and not to form a coiled-coil. Agreement with VirA^ΔDD

OzmJ^ΔDD

M**STAE**FRDILD**RY**ARGAV**GAEQ**ARDLLAARR**FAQ**APGR**PTAP**PSG**QDA**TGPAD**RP**ETL
 3 16 20 29

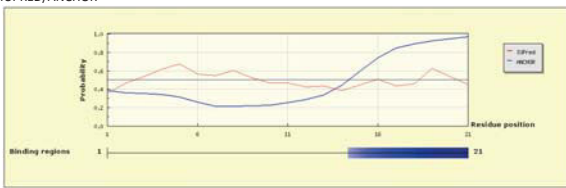


OBSERVATIONS
 Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be borderline disordered (IUPred) to somewhat ordered (PONDR) but not to form a coiled-coil. Agreement with VirFG^ΔDD

OzmK^ΔDD

2 8 12 19
E A E L V E R I E E L P E D T V A T L L A

IUPRED/ANCHOR



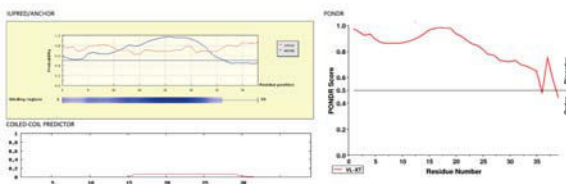
Sequence is too short for PONDR/COILED-COIL PREDICTOR

OBSERVATIONS

The second α -helix predicts to be an interaction region (ANCHOR). The docking region is predicted to be borderline disordered (IUPred).
Some agreement with VirA^ΔDD

Ozml^ΔDD

3 13 16 29
M S R E D L A T S R K A R L S P Q K R A L L E K L S G A A R T S G T T R I P R



OBSERVATIONS

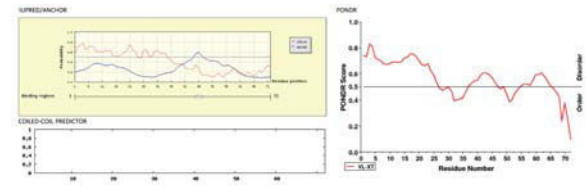
Predicted interaction regions (ANCHOR) overlap well with predicted α -helices. The docking region is predicted to be disordered (IUPred/PONDR), but not to form a coiled-coil.

Strong agreement with VirFG^ΔDD

21

RhiC^ΔDD

39 48 53 60
A D N Q R T A M A V S F N D L D H A V S A G N N V R G Q S R P I E I H P P F A D Q L E G L F L S G E L S I D S L L N I V S P G V A E I R E V S I



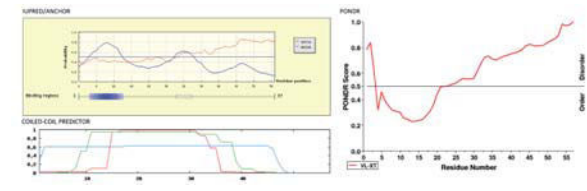
OBSERVATIONS

The secondary structure elements are not predicted to be interaction regions (ANCHOR). The docking region is predicted to be borderline disordered (PONDR) to ordered (IUPred), and not to form a coiled-coil.

Poor agreement with VirA^ΔDD

RhiD^ΔDD

2 12 19 36
(M) S S Q E F Q S I I A G L Q N R Q I T V A E A K Q R L Q R L K A A S A T L G T E Q T E A A D K A V S G T G R A Q E P



OBSERVATIONS

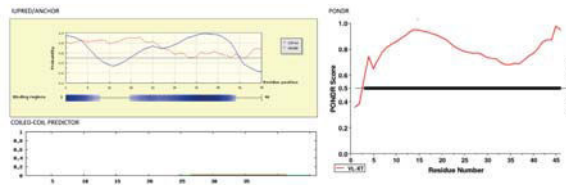
The first predicted α -helix is predicted to be an interaction region, but the second only weakly. The docking region is predicted to be mixed ordered/disordered (IUPred/PONDR) and strongly predicted to form a coiled-coil.

Does not agree with VirFG^ΔDD; may be an alternative type of docking domain?

22

RizC^ΔDD

29 37 40
R A L V T E A P P R G A T Q S A N G H 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



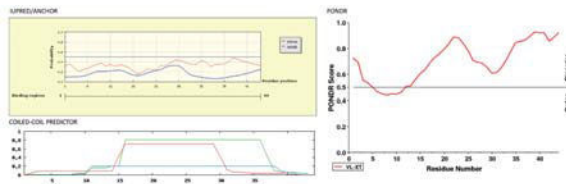
OBSERVATIONS

The predicted α -helices are predicted to be an interaction region (ANCHOR). The docking region is predicted to be borderline (IUPred) to strongly ordered (PONDR), and not to form a coiled-coil.

Agreement with VirA^ΔDD

RizD^ΔDD

7 42
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 A K S A S S E A F A



OBSERVATIONS

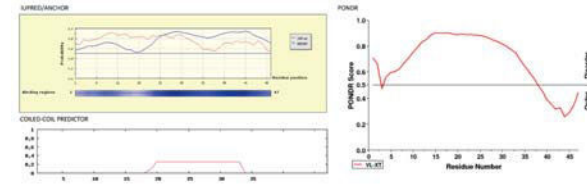
The predicted α -helix is not predicted to be an interaction region (ANCHOR). The docking region is predicted to be borderline disordered (IUPred) to disordered (PONDR), but to form a coiled-coil.

Does not agree with VirFG^ΔDD; may be an alternative type of docking domain?

23

Ta-1^ΔDD

21 29 35 44
G N E I Q T V S A G A P A Q G E P P S L D E L L R Q V E A G E L D P S V A Q Q F L N S Q S



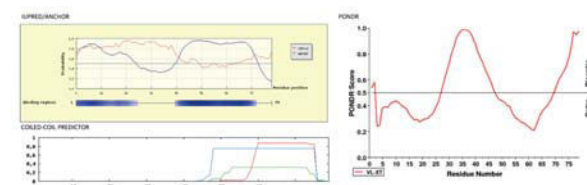
OBSERVATIONS

The predicted α -helices are predicted to be interaction regions (ANCHOR). The docking region is predicted to be disordered (IUPred) to mixed disordered/ordered, and not to form a coiled-coil.

Agreement with VirA^ΔDD

TaO^ΔDD

73 43 55 58
(M) G E G H R I G V L Q Q P P R G R D C H E A D D G S R R I P Q P P V P Q V G R P G R G A V M T R L E I L Q A L Q S G R M A L A D A K K A L A A I E A Q V Q P E D



OBSERVATIONS

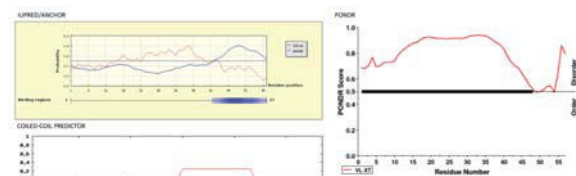
The predicted α -helices are predicted to be interaction regions (ANCHOR). The docking region is predicted to be borderline (IUPred) to mixed ordered/disordered (PONDR), and the second α -helix to form a coiled-coil.

Does not agree with VirFG^ΔDD; may be an alternative type of docking domain?

24

5925 ⁶DD

LAANAFAGSAVAEAAASVAPAVEATEPAAAPLDTLSEIEGLQDDDLAALLGKFEIFRE

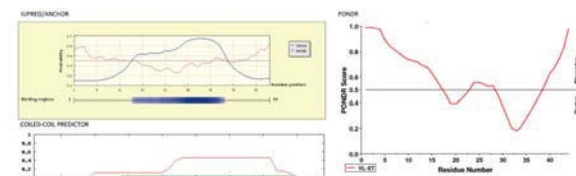


OBSERVATIONS

The second predicted α -helix is predicted to be an interaction region (ANCHOR). The docking region is predicted to be mixed ordered/disordered (IUPred/PONDR), and not to form a coiled-coil. Some agreement with VirA ⁶DD

5924 ⁸DD

MNKP TSSD GWKDDYL SRLSRLSKNQLMALALKLKQQQLQGPAA



OBSERVATIONS

Predicted interaction regions (ANCHOR) overlap well with the predicted α -helix. The docking region is predicted to be borderline disordered to more ordered (IUPred/PONDR), and not strongly predicted to form a coiled-coil.

Agreement with VirG ⁸DD

OBSERVATIONS

This region forms a coiled-coil (correctly predicted by the coiled-coil predictor). It is also correctly predicted (ANCHOR) to form an interaction motif, and by both IUPred and PONDR to be disordered (consistent with its coiled-coil nature).

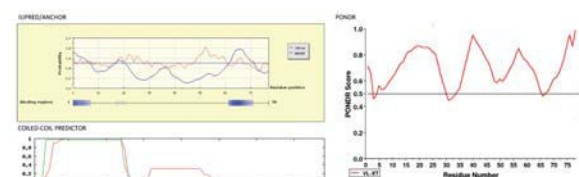
CONTROL: CLASS 1A DDs

Key:

Red = α -helices observed in the NMR structure of the DEBS 2 ⁶DD/DEBS 3 ⁸DD complex
Blue = α -helices predicted by PsiPred

DEBS 2 ⁶DD

AASPAVDGGRMDLEKALALSAEDCHDWDGQRLSLELRVNSRRADAPSTSAISEDASDELFSM
DGRFGGGEDL

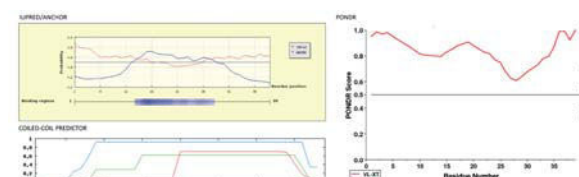


OBSERVATIONS

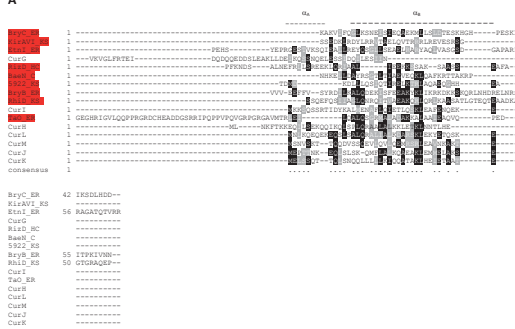
The third docking α -helix is correctly predicted to be interaction region (ANCHOR), and to be relatively disordered (IUPred/PONDR). The first two α -helices, which form a dimerization motif (four α -helical bundle), are predicted to be borderline disordered (IUPred) to disordered (PONDR), and to form a coiled-coil motif.

DEBS 3 ⁸DD

MSGDNGMTBRLRRYLKRTVTELDSTVIALRVEVHRAGE



A



B

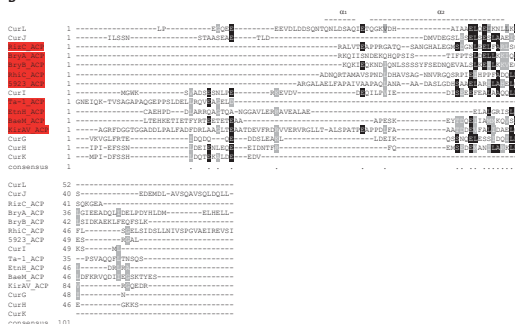


Figure 1. Comparative sequence analysis of docking domains from *trans*-AT PKS and Class 2 docking domains. (A) Multiple alignment using ClustalOmega¹¹ of the N-terminal regions of PKS subunits from *trans*-AT PKS (in red), with those of Class 2 N-terminal docking domains of *cis*-AT PKS.¹² The positions of the two α -helices in the solved structure of the CurH and CurL N-terminal docking domains are indicated, the second of which forms a coiled-coil. (B) Multiple alignment using ClustalOmega¹¹ of the C-terminal regions of PKS subunits partnering those in (A) (in red), with those Class 2 C-terminal docking domains of *cis*-AT. The positions of the two α -helices in the solved structure of the CurG and CurK C-terminal docking domains are indicated. Only Ta-1 does not align particularly well with the others.

ASSOCIATED REFERENCES

- (1) Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I. *Bioinformatics* **2005**, *21*, 3433.
- (2) Dosztányi, Z.; Mészáros, B.; Simon, I. *Bioinformatics* **2009**, *25*, 2745.
- (3) Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. *Proteins* **2001**, *42*, 38.
- (4) Li, X.; Romero, P.; Rani, M.; Dunker, A.K.; Obradovic, Z. *Genome Inform. Workshop Genome Inform.* **1999**, *10*, 30.
- (5) Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. *J. Mol. Biol.* **2004**, *337*, 635.
- (6) Linding, R.; Jensen, L.J.; Diella, F.; Bork, P.; Gibson, T.J.; Russell, R.B. *Structure* **2003**, *11*, 1453.
- (7) Pavlović-Lažetić, G.M.; Mitić, N.S.; Kovačević, J.J.; Obradović, Z.; Malkov, S.N.; Beljanski, M.V. *BMC Bioinformatics* **2011**, *12*, 66.
- (8) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.
- (9) Broadhurst, R.W.; Nietlispach, D.; Wheatcroft, M.P.; Leadlay, P.F.; Weissman, K.J. *Chem. Biol.* **2003**, *10*, 723.
- (10) Buchholz, T.J.; Geders, T.W.; Bartley F.E. 3rd; Reynolds, K.A.; Smith, J.L.; Sherman, D.H. *ACS Chem. Biol.* **2009**, *4*, 41.
- (11) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J. Thompson, J.D.; Higgins, D.G. *Mol. Syst. Biol.* **2011**, *7*, 539.
- (12) Whicher, J.R.; Smaga, S.S.; Hansen, D.A.; Brown, W.C.; Gerwick, W.H.; Sherman, D.H.; Smith, J.L. *Chem. Biol.* **2013**, *20*, 1340.

III. Etude d'une enzyme post-PKS impliquée dans la cyclisation de la lankacidine C

1. Analyse *in-silico* de la LkcE

La LkcE est une enzyme post-PKS intervenant dans la macrocyclisation de la lankacidine C. Il est proposé que cette enzyme soit une amine oxydase incorporant un co-facteur flavinique (Arakawa, *et al.*, 2005). Afin de catalyser la réaction de cyclisation, la LkcE permettrait l'oxydation de l'amine en imine, et catalyserait également la déprotonation du C2 du polykétide, pour activer celui-ci afin de former une nouvelle liaison C-C. Afin de pouvoir identifier plus facilement les acides aminés importants pour la catalyse, nous avons cherché d'éventuels homologues. Nous avons tout d'abord utilisé Psi-Blast afin d'identifier dans les banques de données des protéines homologues à la LkcE. Trois LkcE présentant entre 99% et 91% d'identité de séquence ont été identifiés chez des espèces de *Streptomyces* possédant également le cluster de gènes codant la PKS de la lankacidine (*S. griseofuscus*, *S. cellostaticus*, et *S. yokosukanensis*). Une amine oxydase identifiée chez *Hahella chejuensis* présente également une forte identité de séquence avec la LkcE (63% d'identité). De même, une LkcE présente chez *P. fluorescens*, une bactérie ayant un cluster de gènes similaire à celui produisant la lankacidine C chez *S. rochei*, présente 53% d'identité de séquence. D'autres amines oxydases présentent également une identité de séquence assez importante avec la LkcE (de 47% à 29 % d'identité). Les plus proches homologues identifiés au sein de la PDB sont les monoamines oxydases humaines (MAO) A et B (26% et 27 % d'identité de séquence, respectivement).

Les séquences des plus proches homologues ainsi que des MAO humaines ont ensuite été alignées en utilisant le serveur ClustalW de l'IBCP à Lyon (Combet, *et al.*, 2000), combiné au programme Esript (Robert, et Gouet, 2014) (Figure 72).

Un motif typique de domaine de fixation a été identifié. En effet, un motif consensus GxGxxG(x)₂₀E est présent dans la séquence de la LkcE de *S. rochei* (à partir de la G15.). Ce motif est caractéristique des protéines de la famille des glutarédoxines mais est également présent chez certaines monoamines oxydases (Dym, et Eisenberg, 2001).

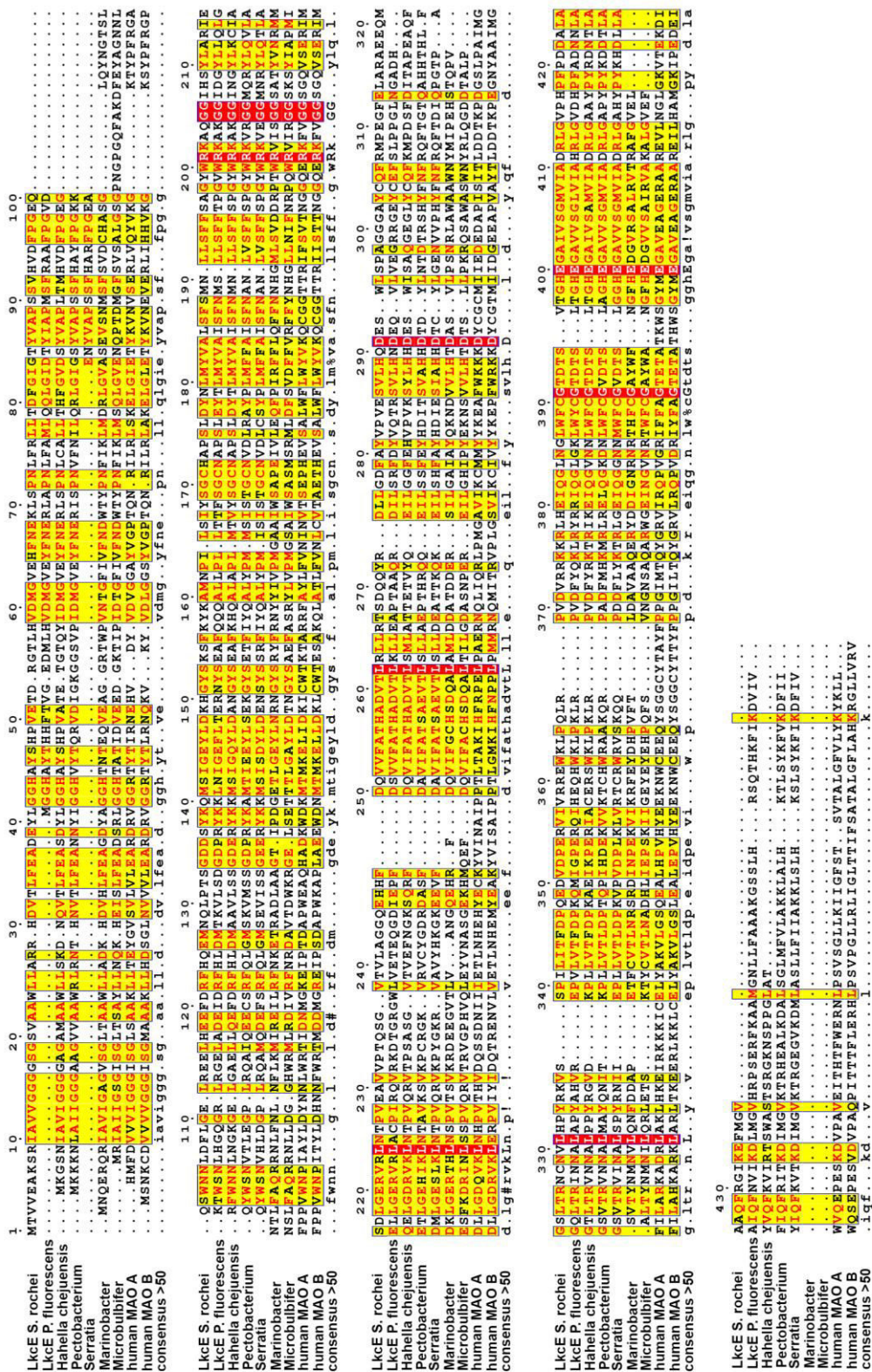


Figure 72 : Alignement des séquences des protéines homologues à la LkcE. La numérotation utilisée est basée sur la LkcE de S. rochei. Les résidus conservés dans l'ensemble des séquences sont indiqués en rouge, et ceux conservés dans minimum 50% des séquences sont indiqués en jaune. Les protéines intégrées dans cet alignement ont été identifiées par Psi-Blast.

2. Clonage et purification de la LkcE et des mutants

A) Clonage de la séquence codant la LkcE

A partir de la souche fournie par l'équipe du Pr P. F. Leadlay qui contient le gène de la LkcE inséré dans le plasmide pGEX4T-3, une mini-préparation pour récupérer le plasmide a été réalisée. Une coupure du vecteur par *Bam*HI et *Xho*I a permis d'exciser la séquence codante, qui a été ensuite sous clonée dans le pBG102. Une transformation de bactérie *E. coli* DH5 α a permis de propager ces vecteurs. Le séquençage des deux plasmides a permis de s'assurer non seulement de la présence de la séquence codante mais aussi de son intégrité.

B) Purification de la LkcE

Un protocole de purification de la LkcE a été mis au point et comporte 6 étapes : i) la lyse des cellules par French Press permettant de récupérer environ 2/3 de l'enzyme sous forme soluble ; ii) une chromatographie de pseudo-affinité utilisant le tag His₆ pour purifier la protéine par chromatographie type chélation de métaux immobilisés (IMAC) sur colonne nickel par une étiquette His₆ ; iii) l'excision par la protéase rhinovirus humain 3C de la protéine de fusion SUMO taguée His₆ pendant dialyse ; iv) une seconde utilisation de l'IMAC pour séparer la LkcE de la protéine SUMO-His₆, la LkcE n'étant alors pas retenue sur la colonne nickel ; v) dilution de la protéine dans un tampon de faible force ionique avant l'utilisation d'une chromatographie échangeuse d'anions Q-sépharose (la LkcE est alors éluée à une concentration de 400 mM de NaCl) ; vi) une chromatographie d'exclusion stérique sur colonne Superdex 200 26/60 (la protéine est éluée à 195 mL). La pureté de la protéine a été suivie à chaque étape par gel SDS-PAGE 12,5%. Ce protocole permet d'obtenir une protéine très pure tout en minimisant le temps de purification (2 jours) (Figure 73). Un rendement moyen de 15 mg/L de culture a pu être obtenu pour les différentes purifications de la LkcE. Un dosage spectrophotométrique en UV-visible a permis d'estimer à environ 45 % le pourcentage de protéine ayant incorporé le FAD.

Ce protocole de purification a également été employé pour purifier chacun des mutants de la LkcE généré plus tard (Y182F, R326L et R326Q).

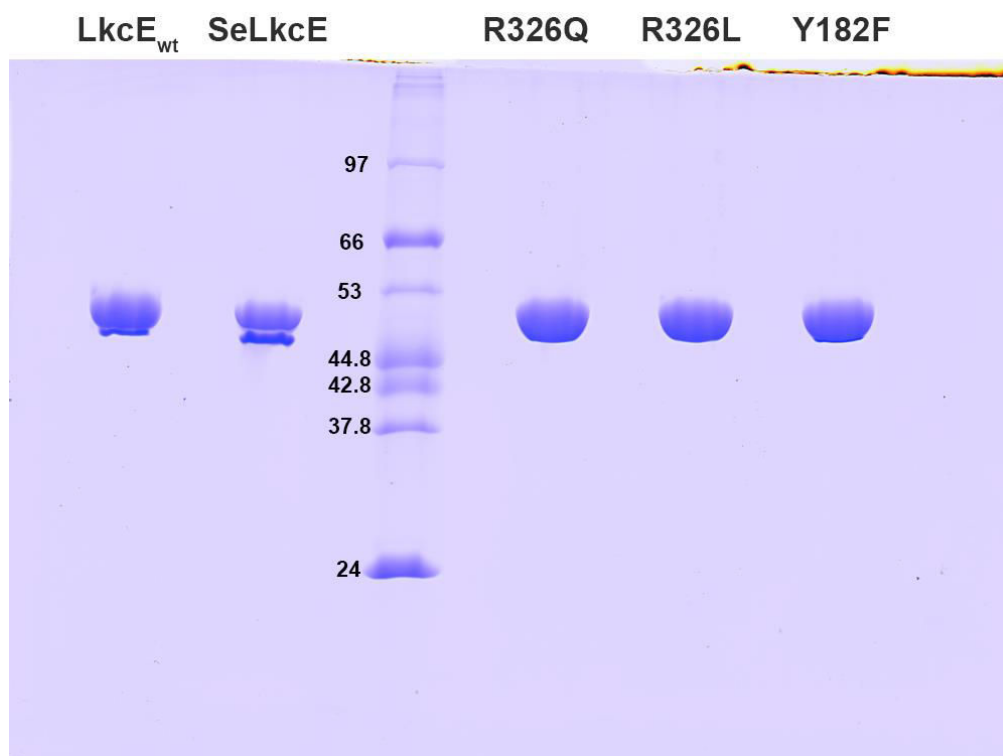


Figure 73 : Analyse par gel SDS-PAGE 12,5% de chaque construction de la LkcE après purification.

Nous pouvons voir que la protéine a été purifiée à homogénéité. Seule une seconde bande, de faible intensité pour la LkcE_{wt} et en plus grande proportion pour la SeLkcE est visible. Celle-ci pourrait correspondre aux formes oxydées et réduites de l'enzyme ou bien à de la protéolyse dans l'échantillon.

C) Homogénéité en solution

Dans le but de mener des études structurales sur la LkcE, nous devons nous assurer que la protéine a bien été purifiée à homogénéité. Pour cela, nous avons utilisé la diffusion dynamique de la lumière (DLS). Il s'agit d'une technique non destructive très sensible aux hétérogénéités en solution et permettant d'estimer la qualité de la protéine purifiée. A l'issue de la purification, la LkcE, ainsi que chacun des mutants, ont été analysés par DLS (Tableau 1).

	Polydispersité	Diamètre hydrodynamique moyen (en nm)	Atténuateur
LkcE _{wt}	0,016	7,977	8
SeLkcE	0,051	7,979	8
LkcE Y182F	0,020	7,668	8
LkcE R326Q	0,076	8,029	9
LkcE R326L	0,101	8,416	9

Tableau 1 : Données DLS obtenues après purification de la LkcE_{wt}, la SeLkcE et des mutants Y182F, R326Q et R326L.

La valeur de l'indice de polydispersité n'excède pas 0,1, ce qui démontre l'homogénéité de la solution protéique. De plus, la valeur du diamètre hydrodynamique moyen reste constante. La valeur d'atténuateur rend compte de la concentration de la solution protéine.

La présence d'un pic unique et fin, ainsi qu'un indice de polydispersité compris entre 1 à 10 % montre que chaque version de LkcE a bien été purifiée à homogénéité et que la protéine n'agrège pas dans son tampon après filtration sur gel. La qualité des échantillons était donc amplement suffisante pour procéder à des études structurales.

D) Vérification de l'intégrité de la LkcE_{wt} et caractérisation de la nature du co-facteur flavinique

Une chromatographie en phase reverse C8 en présence de 0,1% de TFA a permis de séparer l'enzyme de son cofacteur, indiquant que ce dernier n'est pas lié covalamment à l'enzyme. Les deux ont été analysés par spectrométrie de masse en condition dénaturante par ionisation en électrospray (ESI) (F. Dupire, Service Commun de Spectrométrie de Masse de la Faculté des Sciences et Technologie de l'Université de Lorraine). La masse moléculaire obtenue par déconvolution (spectre en mode positif) pour la protéine est 49 289,3 Da (Figure 74A) ce qui correspond à la masse théorique (49 289,4 Da). En ce qui concerne la flavine, une masse moléculaire de 785,1 Da a été obtenue (Figure 74B), ce qui correspond à la masse molaire du FAD (785,5 Da) et non à celle du FMN (456,3 Da). On peut donc affirmer que la LkcE a intégré le FAD comme co-facteur.

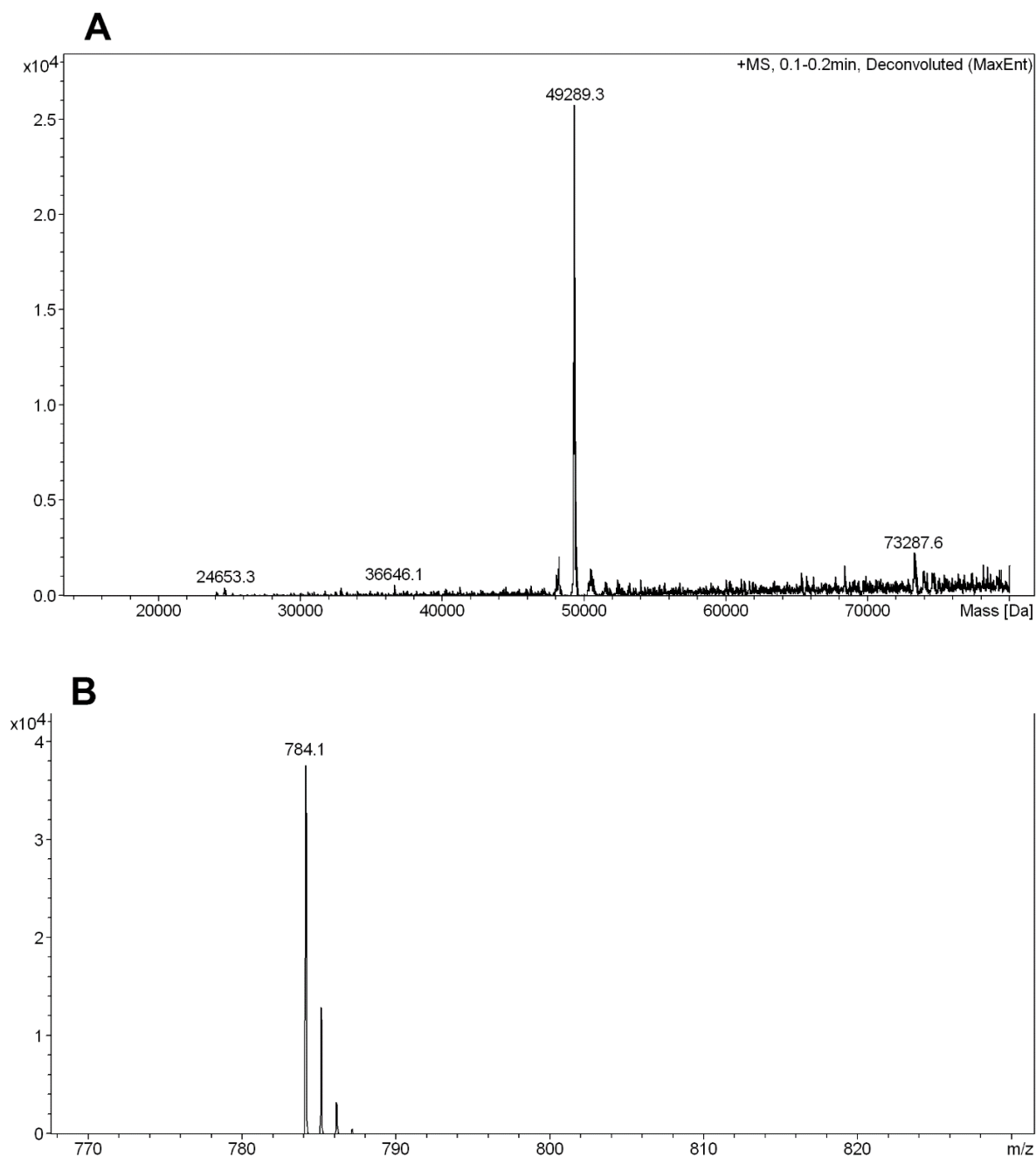


Figure 74 : Détermination de la nature du co-facteur flavinique par spectrométrie de masse (ionisation électrospray).

A : Spectre de masse déconvolué correspondant à la fraction de chromatographie en phase contenant la LkcE.

Une masse moléculaire de 49 289,3 Da a été déterminée (la masse moléculaire théorique de la LkcE étant de 49 289,4 Da).

B : Spectre de masse de la fraction contenant le co-facteur.

On détermine une masse moléculaire de 785,1 Da ($m/z = 784,1$ en mode négatif), correspondant à celle du FAD (785,5 Da).

3. Etude structurale de la LkcE

A) Criblage des conditions de cristallisation de la LkcE_{wt} et cristallisation de la Se-LkcE

Afin d'identifier les conditions permettant d'induire la cristallisation de la LkcE_{wt}, nous avons criblé un grand nombre de conditions de cristallisation à l'aide des *screens* commerciaux Morpheus, Wizard, SaltRX, PGA Screen, PEG/Ion Screen, HR, et Index. Ces *screens* ont été utilisés à deux concentrations en LkcE différentes. A 10 mg/mL, de forts précipités amorphes ont été obtenus dans de très nombreuses gouttes, signifiant que la protéine était trop concentrée et/ou faiblement soluble. Nous avons donc divisé la concentration protéique par deux. Une condition dans le *screen* Index a alors permis l'apparition de cristaux : 25% PEG 3350, 0,2 M d'acétate d'ammonium et 0,1 M Bis Tris (pH 6,5). Ces premiers cristaux se présentaient sous la forme de plaquettes collées les unes aux autres d'environ 150 µm dans la plus grande dimension (Figure 75A). Il est à noter que ces premiers cristaux ont été obtenus lors de mon stage de Master 2 sur une protéine qui n'avait pas eu d'étape de chromatographie d'exclusion stérique.

Nous avons reproduit la condition en boîte *Linbro* en utilisant la technique de la goutte suspendue. Nous avons alors fait varier la concentration de PEG 3350 de 17% à 30% et le pH de 6,0 à 7,0. Des cristaux ont été obtenus pour des concentrations en PEG 3350 allant de 25% à 30%, pour démontrer la reproductibilité de la cristallisation. En revanche, il s'agissait toujours de cristaux multiples (Figure 75B). Ces résultats suffisamment prometteurs nous ont menés à produire de la protéine sélénée (SeLkcE) pour tenter d'améliorer les cristaux d'une part et de phaser d'autre part, puisque le plus proche homologue dans la PDB (1GOS) de LkcE ne présente que 18,7% d'identité de séquence.

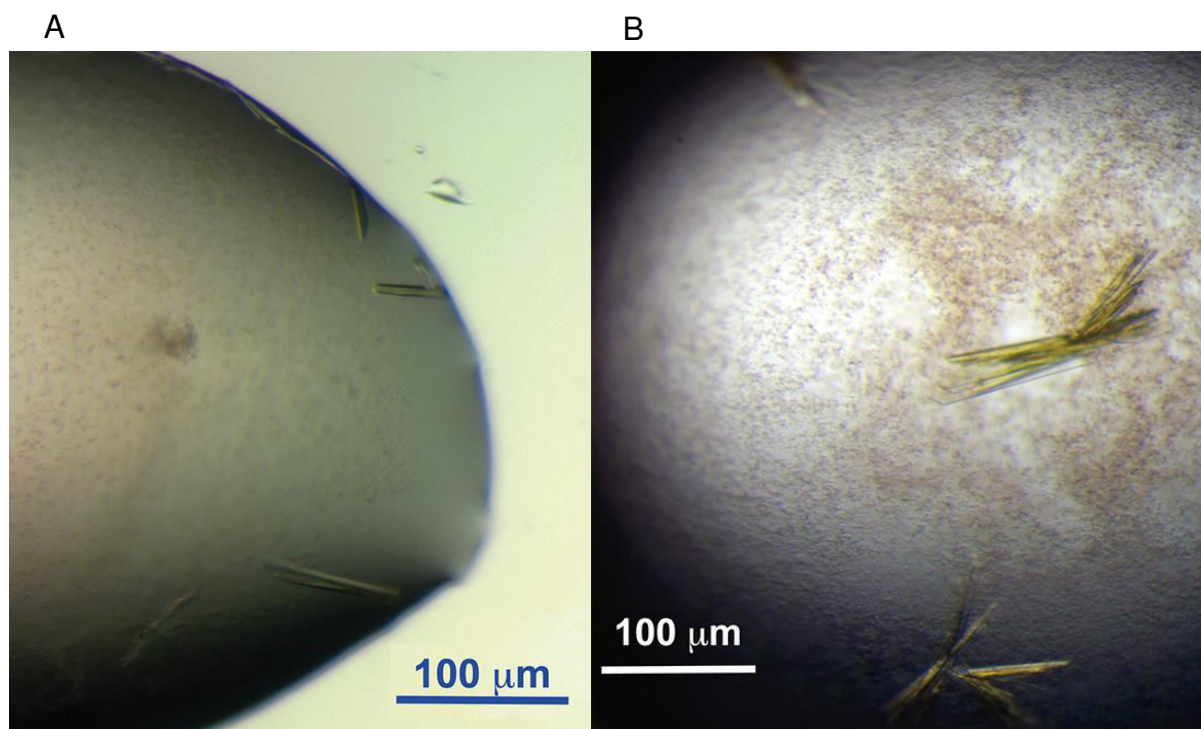


Figure 75 : Premiers cristaux obtenus pour la LkcE.

A : Cristaux en *screen* haut débit de cristallisation. Ces cristaux ont été obtenus en nanogouttes dans le kit de cristallisation Index. La solution de cristallisation permettant l'obtention de ces cristaux contient 25% PEG 3350, 0,2 M d'acétate d'ammonium et 0,1 M Bis Tris (pH 6,5).

B : Cristaux en goutte suspendue. Les cristaux ont été reproduits en boîte *Linbro* en plus grand volume (3 µL). Une gamme de pH (de 6,0 à 7,0) et de concentration en PEG 3350 (de 17,5% à 30%) a permis de comprendre l'effet du pH et de la concentration en agent précipitant. La goutte de la photographie est celle contenant la solution de cristallisation composée de 25% PEG 3350, 0,2 M d'acétate d'ammonium et 0,1 M Bis Tris (pH 6,5).

Des essais de cristallisation avec les *screens* commerciaux ont alors été menés avec la SeLkcE pour à la fois vérifier si celle-ci cristallise dans la même condition que la LkcE_{wt} et si une autre condition de cristallisation conduit à des cristaux moins multiples. Malheureusement, aucun cristal n'a été obtenu en nanogouttes, et ce quel que soit les concentrations de SeLkcE utilisée (5, 8 ou 10 mg/mL).

Nous avons alors fait des gammes en boîte *Linbro* en faisant varier les paramètres susceptibles de cristalliser la protéine sélénée à partir des conditions de cristallisation de la protéine native, à savoir: la concentration en protéine, en agent précipitant et en sel, ainsi que le pH. De la même façon, nous avons fait varier la présence ou non des deux analogues de substrat (EMAA et DATD), ainsi que leur concentration. Nous avons obtenu un seul cristal prismatique, d'environ 150 µm par 100 µm dans la condition suivante : 28% PEG 3350, 200 mM acétate d'ammonium

200 mM, 0,1 M Bis Tris (pH 6,5), SeLkcE (8 mg/mL) et DATD (5 mM) (Figure 76). Ce cristal a donc été obtenu en présence de l'un des deux analogues de substrat, le DATD.

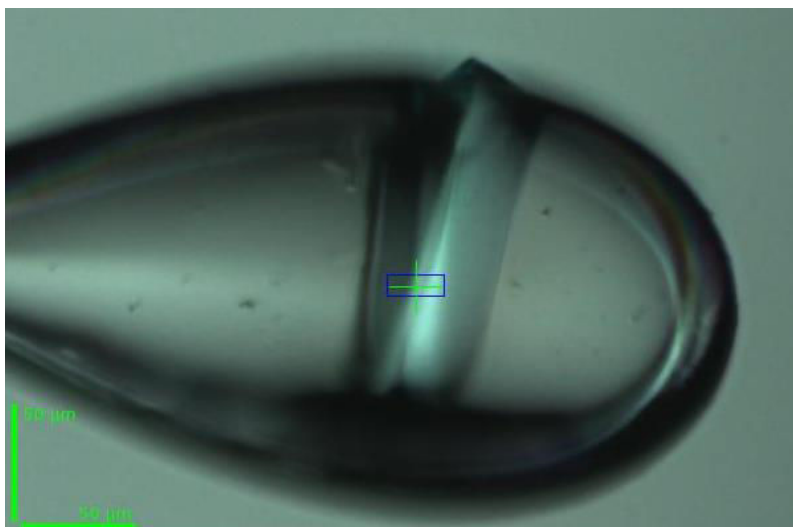


Figure 76 : Photographie du cristal de SeLkcE monté sur une boucle.

Cette photographie a été prise au synchrotron SOLEIL, sur la ligne de lumière Proxima 2. Le carré bleu représente la taille du faisceau de rayon X.

Ce cristal a été congelé en flux d'azote gazeux en utilisant 30% d'éthylène glycol comme cryo-protectant, puis stocké dans l'azote liquide pour la mission synchrotron. Les données de diffraction ont ensuite été enregistrées au synchrotron SOLEIL, sur la ligne Proxima 2.

B) Phasage SAD et construction du premier modèle

Un jeu MAD a été collecté sur la ligne Proxima 2. Pour cela, le seuil d'absorption du diffuseur anormal a toujours d'abord été enregistré (Figure 77).

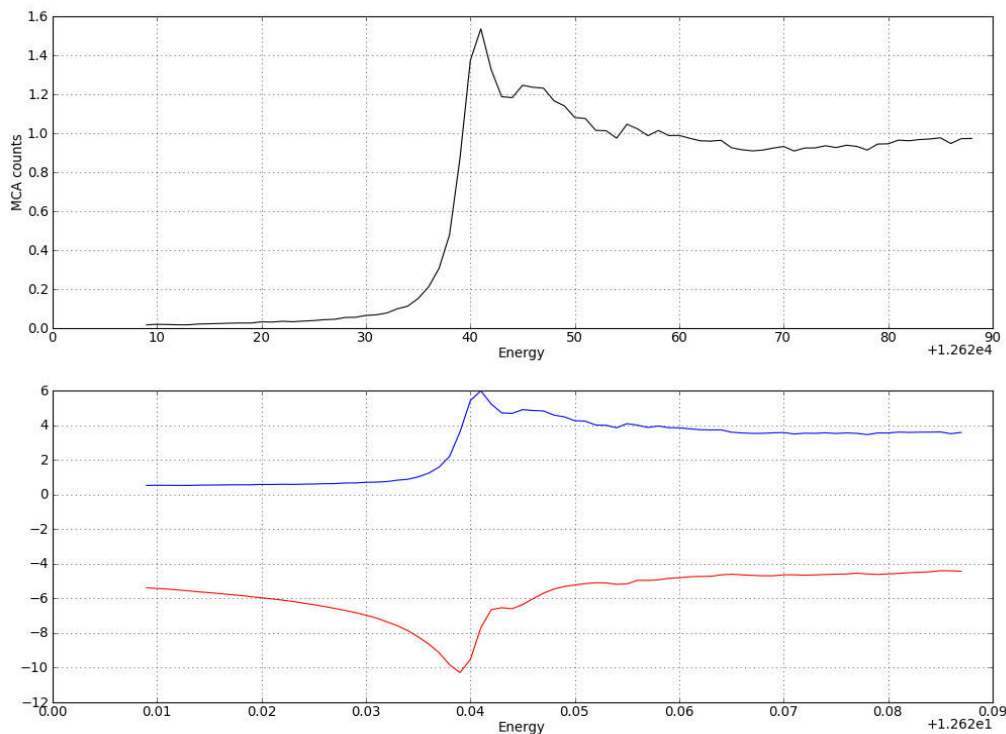


Figure 77 : Mesure du seuil d'absorbance du sélénium.

Le pic d'absorbance se situe à 12,660 keV, ce qui correspond à une longueur d'onde de 0,9793 Å.

Les clichés ont été enregistrés sur 360° avec une oscillation $\Delta\phi$ de 0,5° par image. Les données ont été indexées en utilisant le programme XDS, avec les paramètres de maille 125,1 Å, 125,1 Å, 154,6 Å, 90,0° 90,0° 90,0°. Le groupe d'espace tétragonal a pu être déterminé au vue des extinctions systématiques ; il s'agit du P4₁2₁2 (92). Les données ont ensuite été intégrées avec XDS jusqu'à une résolution de 3,3 Å, puis mises à l'échelle avec le programme XSCALE. Le contenu de la maille a été évalué grâce au calcul du coefficient de Matthews (Matthews, 1968), pour un pourcentage de solvant de 60% avec deux molécules dans l'unité asymétrique.

Les phases ont été déterminées par SAD en utilisant uniquement le jeu de données du *peak*. Pour cela, la suite SHELX-CDE a été utilisée en utilisant les paramètres déterminés grâce au calcul du coefficient de Matthews. La position des atomes de sélénium et la main a été déterminée grâce à SHELX-D. L'affinement des

coordonnées des atomes de Se a été réalisé avec SHELX-E, ainsi que la détermination des premières phases. Les phases ont ensuite été améliorées par aplatissage de solvant avec le programme PARROT. La carte de densité obtenue à 3,3 Å de résolution était de qualité suffisante pour identifier les éléments de structure secondaire dans la protéine.

Cette source de phase, a permis de construire un premier modèle de la structure en utilisant le programme de construction automatique Buccaneer. Celui-ci est particulièrement efficace pour construire dans des cartes de densité électronique à basse résolution (>2,5 Å). Buccaneer est parvenu à construire 76% de la protéine, soit 672 résidus sur les 884 acides aminés composant la séquence. Cependant le programme commet de nombreuses erreurs qu'il a fallu corriger, et il a fallu en plus de construire les résidus manquants. Après plusieurs cycles d'affinement, les statistiques sont présentées dans le Tableau 2.

Groupe d'espace	P4 ₁ 2 ₁ 2 (92)
Paramètre de maille (Å)	125,1 125,1 154,6 90,0 90,0 90,0
Ligne de lumière	Proxima II
Longueur d'onde	0,9793
Résolution	20-3,3 Å
Complétude	100% (100%)
Réflexions observées	210 559 (46 298)
Réflexions uniques	49 383 (10 724)
Redondance	4,26 (4,32)
I/σI	10,63 (2,36)
Signal anomal	1,523 (0,728)
R _{merge} (%)	13,8 (71,8)
Nombre de molécules dans l'unité asymétrique	2 (1 dimère)
Pourcentage d'acides aminés modélisés	80 % (362 résidus)
R factor	0,33
R _{free}	0,37

Tableau 2 : Statistiques des données de diffraction et d'affinement de la structure de la SeLkcE.

Un certain nombre de résidus sont absent de la densité électronique dans des boucles trop flexibles pour pouvoir être construites.

C) Cristallisation de la LkcE avec analogues de substrat et construction du modèle

a. Obtention de cristaux monocliniques

Pour améliorer la qualité de la densité électronique, de nombreux essais de cristallisation ont été menés pour obtenir des cristaux qui diffractent mieux que la protéine sélénée. En parallèle, nous avons donc affiné les conditions de cristallisation de la LkcE, en présence des deux analogues de substrat (EMAA et DATD à 10 mM). Ces cristaux ont été obtenus en changeant le contre ion de l'ammonium. Ainsi l'acétate d'ammonium a été remplacé par du formate d'ammonium. Des cristaux, peu reproductibles, sous forme de plaquettes plus ou moins épaisses, ont pu être obtenus dans la condition suivante : 24 à 26% PEG 3350, 400 mM formate d'ammonium, 100 mM Bis Tris (pH 6,4 à 6,6), 2,5 mM Tris(2-carboxyethyl)phosphine (TCEP) en utilisant une concentration en protéine de 5 à 7 mg/mL (Figure 78). Les cristaux ont été congelés en flux d'azote, en utilisant 30% d'éthylène glycol comme cryo-protectant.

Des données de diffraction ont été enregistrées jusque 2,65 Å sur la ligne de lumière Proxima 1. Une analyse du contenu de la maille grâce au coefficient de Matthews laissait présager la présence de 4 dimères dans l'unité asymétrique avec 45% de solvant, ou bien 3 dimères dans l'unité asymétrique avec 59% de solvant.

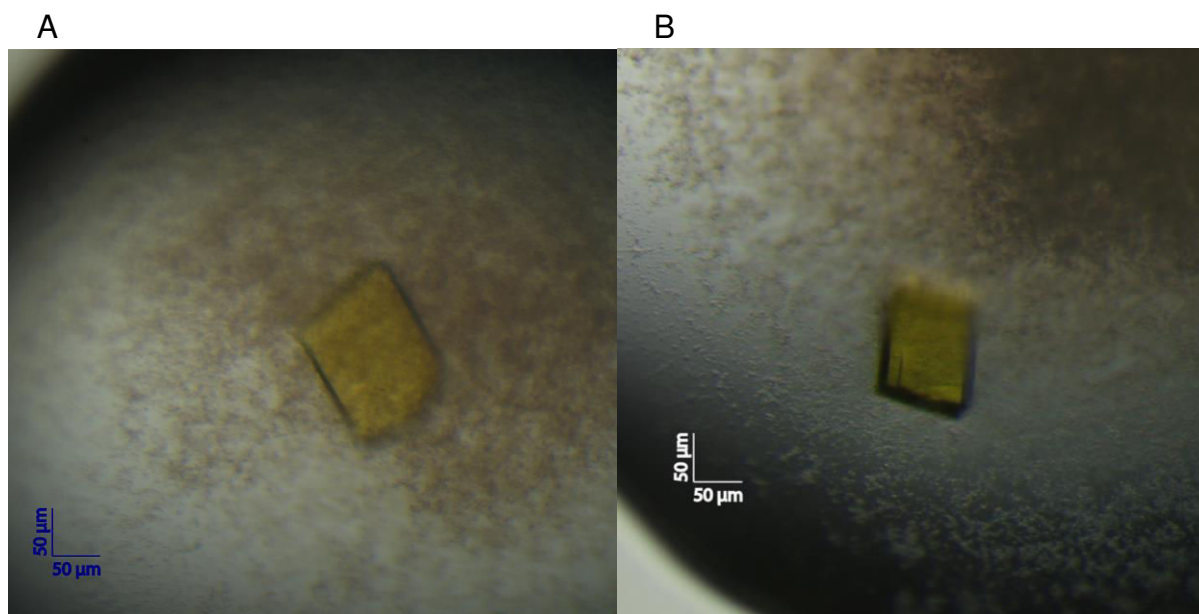


Figure 78 : Photographie des cristaux de LkcE en présence des deux analogues de substrat.

Ces cristaux sont sous la forme de plaquette, ici relativement épaisse, d'environ 150 µm/100 µm obtenus respectivement dans les conditions : (A) 24% PEG 3350 et (B) 26% PEG 3350, 400 mM formate d'ammonium, 100 mM Bis Tris (pH 6,5), 2,5 mM TCEP.

Les données ont été phasées par remplacement moléculaire en utilisant le programme Phaser MR avec pour modèle la structure de la SeLkcE. Par remplacement moléculaire, la solution avec 8 molécules (4 dimères) dans l'unité asymétrique présente les meilleures statistiques. En effet, Phaser MR n'évalue pas les solutions proposées en termes de facteur R, mais par un score qui rend compte de la probabilité des solutions. Seules les solutions dont le Z-score est supérieur à 6 sont conservées, les autres étant jugées improbables. Pour la fonction de rotation, huit solutions avec des valeurs de Z-score de 14,2 à 23,49 ont été conservées. La fonction de translation était tout aussi contrastée, et huit solutions de Z-score compris entre 35,8 et 44,0 ont été retenues. Ceci a permis d'obtenir une carte de densité parfaitement interprétable et en adéquation avec la structure de la SeLkcE utilisée comme modèle pour le remplacement moléculaire, attestant également de la réussite du phasage.

Suite au phasage, la construction a été effectuée manuellement avec le programme Coot, en prenant comme base, pour chacune des 8 chaînes, le modèle construit pour la SeLkcE. Ce jeu de données, à plus haute résolution (2,65 Å), a permis de construire les résidus manquants sur le modèle précédent. En revanche, la boucle

132–141 n'a pas pu être construite dans aucune des 8 chaînes, ainsi que la boucle 235–240 dans certaines chaînes peptidiques. De plus, bien que la LkcE ait été co-cristallisée avec 10 mM de chacun des deux analogues de substrat, aucune densité électronique interprétable n'est présente au sein du site actif de l'enzyme. Nous n'avons donc pas pu positionner les analogues de substrat.

A ce stade l'affinement les statistiques du jeu de données sont les suivantes (Tableau 3) :

Collecte du jeu de données	
Groupe d'espace	P2 ₁ (4)
Dimensions de la maille	
a, b, c (Å)	136,55 ; 68,8 ; 190,03
α , β , γ (°)	90,0 ; 96,5 ; 90,0
Résolution (Å)	48-2,65 (2,8-2,65)
R _{merge}	0,088 (0,579)
I/ σ (I)	13,61 (2,80)
Complétude (%)	99,8 (99,8)
Réflexions observées	387 798 (58 706)
Réflexions uniques	102 500 (15 429)
Redondance	3,8 (3,8)
Affinement	
Résolution (Å)	48-2,65
Nombre de réflexions	
Working set	102499
Test set	5116
R _{work} /R _{free}	0,264/0,303
Déviations quadratiques moyennes	
Longueurs de liaison (Å)	0,008
Angles de liaisons (°)	1,12

Tableau 3 : Statistiques des données de diffraction et d'affinement de la structure de la LkcE_{wt}.

b. Obtention de cristaux tétraogonaux pour la protéine native

Dans l'espoir d'obtenir des données à plus haute résolution et moins de molécules à affiner dans l'unité asymétrique, nous avons tenté de reproduire le cristal tétraogonal. En parallèle, la purification de la LkcE a été optimisée par l'ajout d'une étape de chromatographie par exclusion de taille. Alors que la protéine ainsi purifiée ne cristallisait pas dans les conditions précédentes, l'utilisation des *screens* commerciaux a permis d'identifier une condition dans le *screen* JCSG+ (14,4% PEG 8000, 20% glycérol, 160 mM acétate de calcium, 80 mM cacodylate de sodium (pH 6,5)) qui a permis l'obtention de cristaux. Ces cristaux très prismatiques (Figure 79) étaient bien plus prometteurs que les cristaux obtenus jusqu'alors et laissaient présager un changement de groupe d'espace.

En effet, nous avons pu obtenir des cristaux reproductibles dans 13 à 17 % de PEG 8000, cacodylate de sodium ou BES pH 6,3 à 6,7, pour des ratios protéine/solution de cristallisation 2/1 et 3/1. Une concentration en LkcE de 5 à 8 mg/mL a permis l'apparition de cristaux, avec une concentration optimale de 6 mg/mL. Ces cristaux ont été également obtenus co-cristallisant avec les deux analogues de substrat.

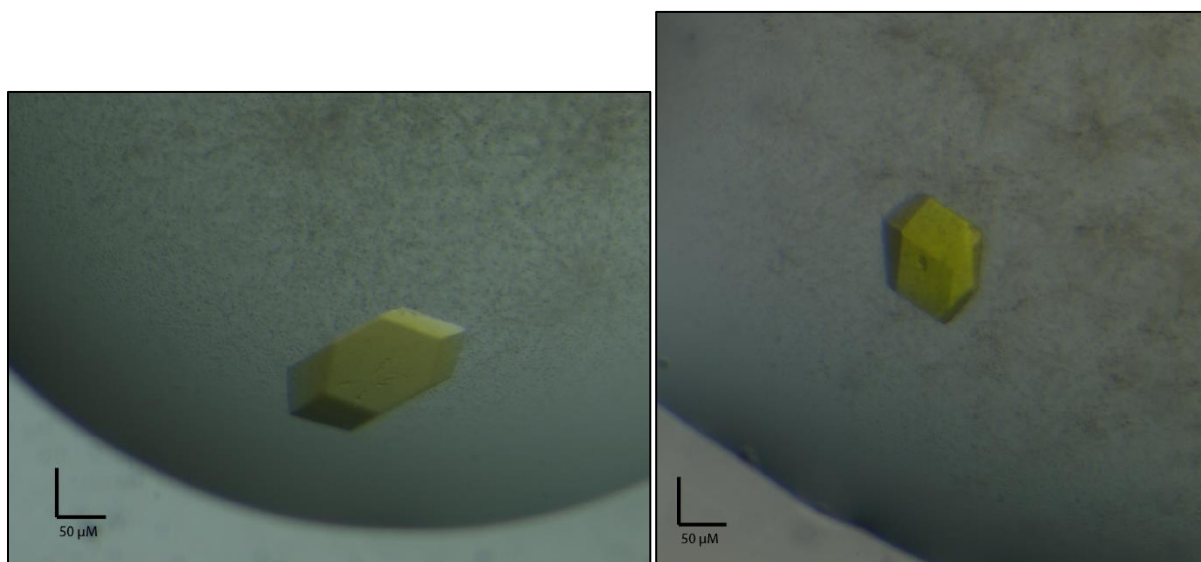


Figure 79 : Photographie des cristaux de LkcE.

Ces cristaux ont été obtenus en utilisant la solution de cristallisation suivante : PEG 8000 (13 à 17%), 160 mM acétate de calcium, 20% glycérol, 80 mM cacodylate de sodium (pH 6,3 à 6,7).

Des données de diffraction ont été enregistrées sur la ligne de lumière Proxima 1 jusqu'à une résolution de 2,8 Å. Après indexation, nous avons pu déterminer qu'il s'agissait du groupe d'espace $P4_12_12$ (92). Il s'agit donc du même groupe d'espace et de la même maille que le cristal de SeLkcE. Une analyse du contenu de la maille a été réalisée en calculant le coefficient de Matthews. Celle-ci est encore une fois identique au cristal de la protéine sélénée puisque l'unité asymétrique est composée de 2 molécules et de 60% de solvant. La phase a été résolue par remplacement moléculaire avec le programme Phaser MR, en utilisant cette fois la chaîne la plus complète construite dans le jeu de données à 2,65 Å de résolution. Pour la fonction de rotation, trois solutions ont été conservées, dont deux ayant un Z-score supérieur à 7. La fonction de translation est quant à elle davantage contrastée. Deux solutions de Z-score 13,78 et 8,13 ont initialement été retenue puis seule la première solution a été conservée pour phaser.

Plusieurs cycles de construction manuelle suivis d'affinement grâce au programme Buster ont été effectués, permettant de mettre en évidence une densité électronique dans le site actif de chaque monomère. Elles ont été attribuées au DATD pour la chaîne A et à l'EMAA pour la chaîne B. L'ambiguïté quant à la présence et au positionnement de ces deux analogues a été levée grâce à la réalisation d'une *omit-map*. Cette dernière a été réalisée en supprimant les deux analogues du fichier PDB et en lançant un cycle d'affinement à l'aide de Buster. Ainsi, si la densité électronique perdure, c'est qu'elle n'a pas été introduite par un biais du modèle. On note que la densité électronique correspondant à l'EMAA au sein du site actif du monomère B est moins bien définie que celle correspondant au DATD dans le site actif du monomère A. En effet, elle apparaît continue à une valeur de ratio signal sur bruit (I/σ_I) faible (carte $2F_o-F_c$ contourée à $0,8 \sigma$).

Les statistiques du jeu de données après affinement sont les suivantes (Tableau 4) :

Collecte du jeu de données	
Groupe d'espace	P4 ₁ 2 ₁ 2 (92)
Dimensions de la maille	
a, b, c (Å)	125.55, 125.55, 156.65
α , β , γ (°)	90.0, 90.0, 90.0
Résolution (Å)	49-2.8 (2.9-2.8)
R _{merge}	0.051 (0.563)
I/ σ (I)	22.37 (3.82)
Complétude (%)	99.9 (100)
Réflexions observées	257 311 (35 009)
Réflexions uniques	26 112 (3 542)
Redondance	7.35 (10.03)
Affinement	
Résolution (Å)	49-2.80
Nombre de réflexions	
Working set	35 112
Test set	1 848
R _{work} /R _{free}	0.192/0.222
Déviations quadratiques moyennes	
Longueurs de liaison (Å)	0.010
Angles de liaisons (°)	1.20

Tableau 4 : Statistiques des données de diffraction et d'affinement de la structure de la LkcE_{wt} en complexe avec les deux analogues de substrat.

Le DATD est positionné au sein du site actif de l'enzyme, à proximité du FAD. Il est principalement maintenu dans le site actif de l'enzyme par des interactions hydrophobes avec la Y168, la Y182, la L185, la L364 et la V396, le groupement méthyle de la T397 et le FAD (Figure 80). Une liaison hydrogène est également possible entre la T397 et la fonction amine du DATD. L'EMAA est également stabilisé au sein du site actif du monomère B par des interactions de type hydrophobe (L70, Y168, Y182, L191, W200, T397) (Figure 81). En revanche, la position de l'EMAA et du DATD au sein du site actif de chaque monomère se superpose.

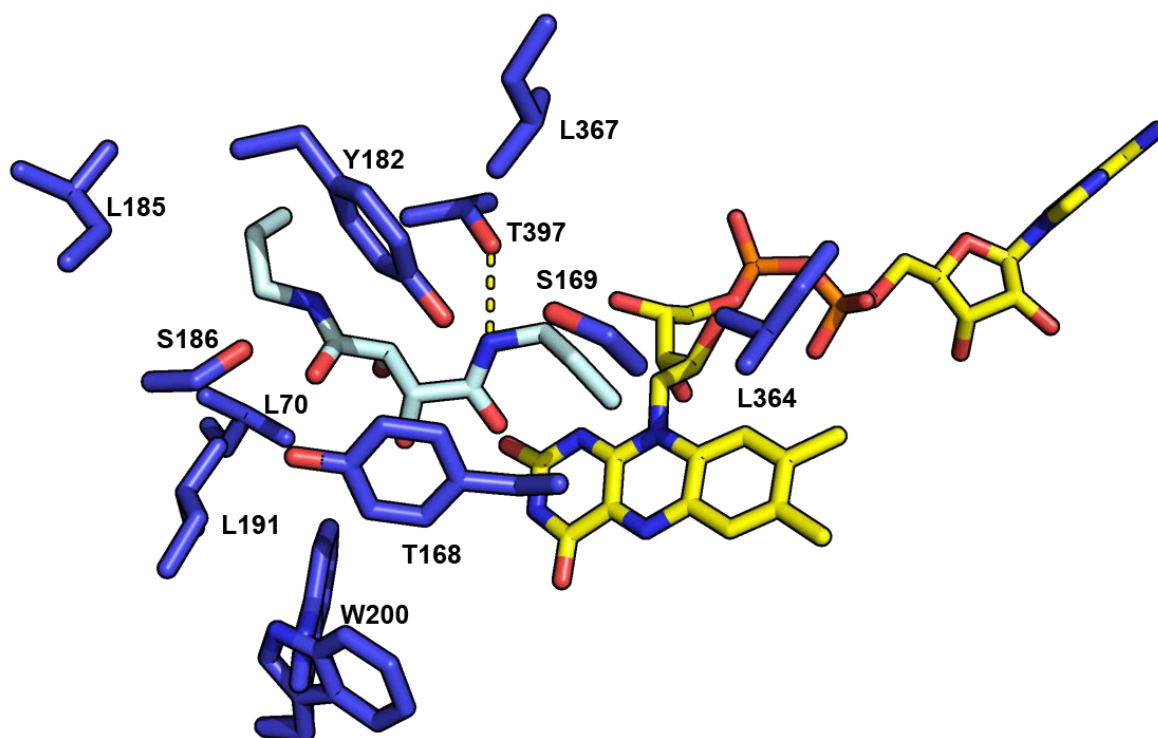


Figure 80 : Vue du DATD au sein du site actif du monomère A de la LkcE.

Seuls les résidus ayant un atome à moins de 5 Å du DATD sont représentés et les résidus pouvant être impliqués dans la stabilisation du DATD sont numérotés. Le FAD est représenté en jaune, le DATD en cyan et la LkcE en bleu.

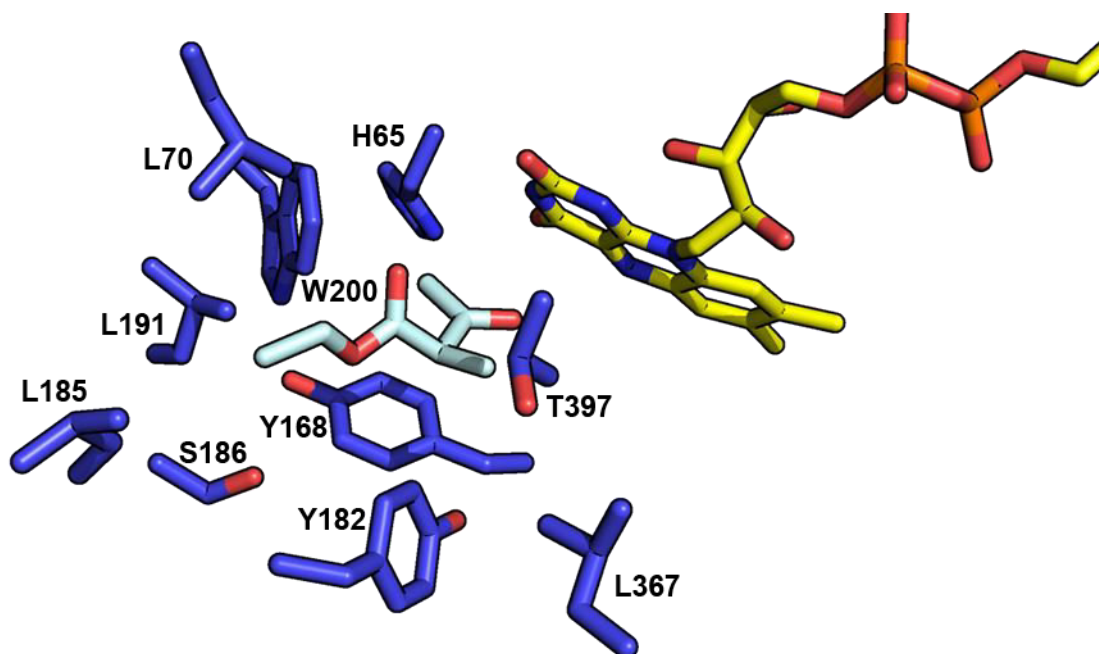


Figure 81 : Site de fixation de l'EMAA.

Seuls les résidus ayant un atome à moins de 5 Å du DATD sont représentés et les résidus pouvant être impliqués dans la stabilisation du DATD sont numérotés. Le FAD est représenté en jaune, le DATD en cyan et la LkcE en bleu.

Sur la base de la structure de la LkcE en P2₁, nous avons identifiés des résidus du site actif pouvant être impliqué dans la catalyse. En effet, la catalyse nécessiterait en principe la présence d'un catalyseur base (voir introduction Figure 51). Il y aurait également besoin de la présence d'un autre résidu pour stabiliser l'imine formée, voire d'un troisième résidu pour stabiliser l'énolate produit durant la catalyse. Sur la base des structures, nous avons donc décidé de produire les mutants Y182F, la tyrosine pouvant jouer le catalyseur base, R326L et R326Q. Les mutants de la R326 ont été réalisés dès la résolution de la structure de la SeLkcE. Il a été surprenant de voir les analogues de substrat être fixé dans le site actif à une distance de la R326 trop grande pour qu'elle puisse jouer un rôle dans leur fixation. Cependant, ce résidu pourrait tout de même être impliqué soit dans la fixation du substrat soit dans la stabilisation de l'énolate formé durant la catalyse. Les mutants de ce résidu ont continué à être utilisés pour la suite du projet. Le but était de rendre la LkcE inactive afin de pouvoir espérer co-cristalliser le substrat dans le site actif de l'enzyme.

D) Caractérisation structurale du mutant R326Q de la LkcE trempé avec le LC-KA05

a. Cristallisation des mutants, collecte de données et affinement de la structure du mutant R326Q en complexe avec le substrat natif

Des gammes de cristallisation ont été menées sur chacun des trois mutants. Ces gammes ont permis d'obtenir des cristaux du mutant Y182F. Ces cristaux ont été obtenus dans les conditions suivantes : 16 à 18 % PEG 8000, 160 mM acétate de calcium, 20% glycérol, 80 mM cacodylate de sodium, pH 6,5 à 6,7.

En revanche, seuls des cristaux de mauvaise qualité ont été obtenus pour les mutants R326L et R326Q. Nous avons alors entrepris d'obtenir des cristaux par *micro-seeding* en utilisant des germes cristallins provenant des cristaux broyés obtenus pour la LkcE_{wt}. Ces cristaux étaient très prismatiques et d'une taille de 100 µm à 300 µm. Dès lors, les cristaux de LkcE_{wt} et du mutant Y182F ont été obtenus de la même manière. La solution de stabilisation des *seeds* est composé de 15% PEG 8000, 160 mM acétate de calcium, 20% glycérol, 80 mM cacodylate de sodium, pH 6,5 ou pH 6,6.

La solution de cristallisation contenue dans les puits est composée ou non de PEG 8000 (0% à 5%), 160 mM acétate de calcium, 20% glycérol, 80 mM cacodylate de sodium (pH 6,5).

Suite à l'extraction du milieu de culture permettant la croissance de la souche de *S. rochei* délétée pour la LkcE (voir section matériel et méthodes), seule une faible quantité de substrat a pu être purifiée par le laboratoire de chimie allemand du Pr A. Kirschning (2 fois 8 mg). Ils ont également pu mettre en évidence une dégradation du LC-KA05. En effet, un départ du groupement acétate a pu être constaté. Celui-ci peut être dû à deux réactions possibles (Figure 82).

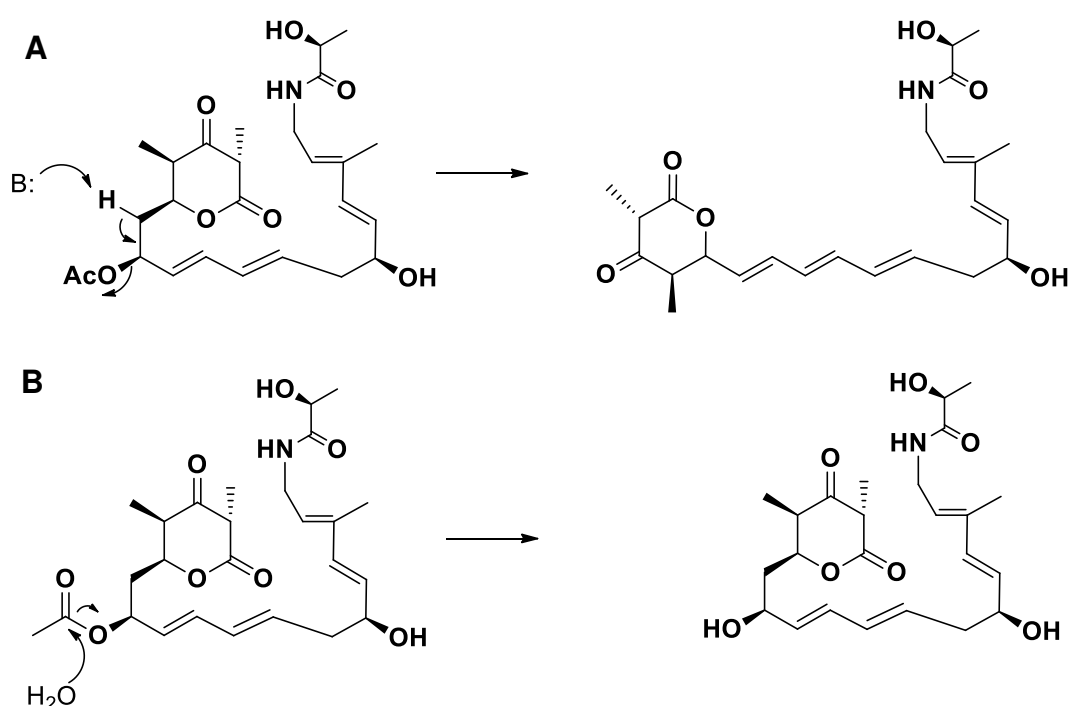


Figure 82 : Réactions de dégradation du LC-KA05.

A : Une base déprotonne le carbone adjacent à l'acétate, ce qui conduit à l'élimination du groupement acétate sous forme de base conjuguée et la formation d'une double liaison C6-C7.
B : *Deuxième mécanisme dans lequel le LC-KA05 est désacétylé par une molécule d'eau.* Il en résulte la formation d'acide acétique et d'un hydroxyle au niveau du C7, qui est plus stable que le groupement acétate.

Afin d'économiser le substrat en vue d'études cinétiques, et pour éviter sa dégradation dans la condition de cristallisation, aucun essai de co-cristallisation n'a été mené. Seuls des trempages des cristaux de l'enzyme sauvage et des trois mutants avec le LC-KA05 ont été entrepris. Les cristaux ont été trempés dans une solution stabilisante à laquelle a été ajouté 30 mM de LC-KA05, dissous dans l'éthylène glycol

(ce dernier jouant le rôle de cryo-protecteur). Les cristaux ont été trempés sur des temps assez courts, de 5 s à 20 minutes.

Les données de diffractions ont été enregistrées sur la ligne Proxima1 et plusieurs jeux de données ont pu être collectés sur chacun des mutants. Ces données ont été phasées avec Phaser MR et les cartes de densité ont été calculées pour chaque jeu de données. Les cristaux du mutant R326Q ont résisté à des temps de trempage assez longs (jusque 20 minutes). Pour les jeux de données avec des trempages de 10 minutes et de 20 minutes, une densité électronique est présente et interprétable comme étant le substrat de l'enzyme au sein du site actif. Seul le mutant R326Q a permis l'observation du substrat dans le site actif de l'enzyme. Le cristal trempé durant 20 minutes a diffracté à 2,9 Å de résolution, alors que celui trempé durant 10 minutes a diffracté à 2,5 Å de résolution. Nous avons donc utilisé ce dernier jeu de données pour la construction et l'affinement.

Nous avons utilisé la structure de la LkcE en complexe avec les analogues de substrat pour phaser en utilisant Phaser MR. La fonction de rotation présentait des Z-score peu élevés mais bien contrastés (deux solutions à 6,25 et 6,48 alors que les autres solutions sont inférieures à 4,5) et la fonction de translation était également très contrastées avec des valeurs de Z-score pour les deux solutions initialement retenues de 14,06 et 13,42.

Le processus d'affinement de ce jeu de données a été effectué à l'aide du logiciel Buster. Seuls quelques cycles d'affinement sont encore nécessaires afin de positionner correctement les dernières molécules de solvant. A ce stade, les statistiques étaient les suivantes (Tableau 5) :

Collecte du jeu de données	
Groupe d'espace	P4 ₁ 2 ₁ 2 (92)
Dimensions de la maille	
a, b, c (Å)	125.55, 125.55, 156.65
α , β , γ (°)	90.0, 90.0, 90.0
Résolution (Å)	49-2.50 (2.65-2.5)
R _{merge}	0.059 (0.686)
I/ σ (I)	17.79 (2.26)
Complétude (%)	99.4 (98.6)
Réflexions observées	192 950 (30 692)
Réflexions uniques	43 804 (6902)
Redondance	4.4
Affinement	
Résolution (Å)	49-2.50
Nombre de réflexions	
Working set	41586
Test set	2189
R _{work} /R _{free}	0.213/0.249
Déviations quadratiques moyennes	
Longueurs de liaison (Å)	0.010
Angles de liaisons (°)	1.17

Tableau 5 : Statistiques des données de diffraction et d'affinement de la structure du mutant R326Q en complexe avec le substrat.

La qualité de la structure a été vérifiée grâce à MolProbity (Figure 83). Les paramètres géométriques ne sont pas parfaits et doivent être légèrement améliorés avant d'envisager un dépôt sur la PDB.

All-Atom Contacts	Clashscore, all atoms:	4.11		99 th percentile* (N=271, 2.50Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	30	4.14%	Goal: <0.3%
	Favored rotamers	634	87.45%	Goal: >98%
	Ramachandran outliers	2	0.24%	Goal: <0.05%
	Ramachandran favored	821	96.47%	Goal: >98%
	MolProbity score [^]	1.89		97 th percentile* (N=6960, 2.50Å ± 0.25Å)
	Cβ deviations >0.25Å	2	0.26%	Goal: 0
	Bad bonds:	14 / 7022	0.20%	Goal: 0%
Bad angles:	54 / 9535	0.57%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 44	0.00%	Expected: ≤1 per chain, or ≤5%

Figure 83 : Résultat de l'analyse de la structure du mutant R326Q par MolProbity.

Le *MolProbity score* combine le *clashscore*, l'évaluation du Ramachandran, l'analyse des rotamères, en un score qui est normalisé par la résolution du jeu de données de diffraction. Le pourcentage est donné par rapport aux autres structures de résolution comparables (100% est le meilleur, 0% le moins bon).

b. Analyse de la structure

La LkcE s'organise en dimère avec une surface d'interaction importante de 1273 Å² entre les deux monomères (calculée grâce au logiciel PISA (Krissinel, et Henrick, 2007)). Cette interface implique 34 à 38 résidus de chaque monomère. Chaque chaîne peptidique est composée de 438 résidus, dont 431 ont pu être construits pour le monomère A et 428 pour le monomère B. En effet, les deux résidus en N-terminale de la chaîne A et les résidus 133–138 situés dans une boucle flexible, n'ont pas pu être construits, de même que les trois premiers acides aminés et les résidus 132–139 de la chaîne B. La LkcE comporte deux domaines : un domaine de fixation du FAD partageant de fortes similarités avec d'autres domaines analogues des monoamines oxydases et polyamines oxydases (Binda, *et al.*, 2001), et un domaine catalytique. Le domaine de fixation diffère d'un domaine Rossmann *fold* par la topologie des éléments de structures secondaires. Le domaine est formé par les résidus 1–45, 70–83, 206–279 et 368–438. Les résidus composant ce domaine de fixation du FAD sont davantage conservés que ceux composant le domaine catalytique (Figure 72). Le domaine est constitué d'un feuillet β central composé de 4 brins parallèles qui est flanqué de 8 hélices α et d'un autre feuillet β composé de 3 brins antiparallèles. Ce domaine sert à l'interaction avec le FAD, et notamment la partie adénosine du FAD (Figure 84). Le serveur DALI (Holm, et Rosenström, 2010) a permis d'identifier des repliements similaires à celui de la LkcE. Ainsi cette dernière présente un repliement assez caractéristique des amines oxydases. Les structures les plus proches sont celles d'une PUFA (*polyunsaturated fatty acid*) isomérase de *Propionibacterium acnes*

(PDB : 2BAB) (Liavonchanka, *et al.*, 2006), d'une protoporphyrinogène oxidase de *Bacillus subtilis* (3I6D) (Qin, *et al.*, 2010), d'une nicotine oxydase de *Paenarthrobacter nicotinovorans* (3NG7) (Kachalova, *et al.*, 2011) et la MAO A humaine (2BXR) (De Colibus, *et al.*, 2005). Bien que le repliement soit sensiblement identique, un alignement de structure deux à deux sur les C α en utilisant ProSMART (Nicholls, *et al.*, 2012) a permis de déterminer des RMSD de 10,7 ; 7,5 ; 10,1 et 13,5 Å respectivement sur l'ensemble des C α .

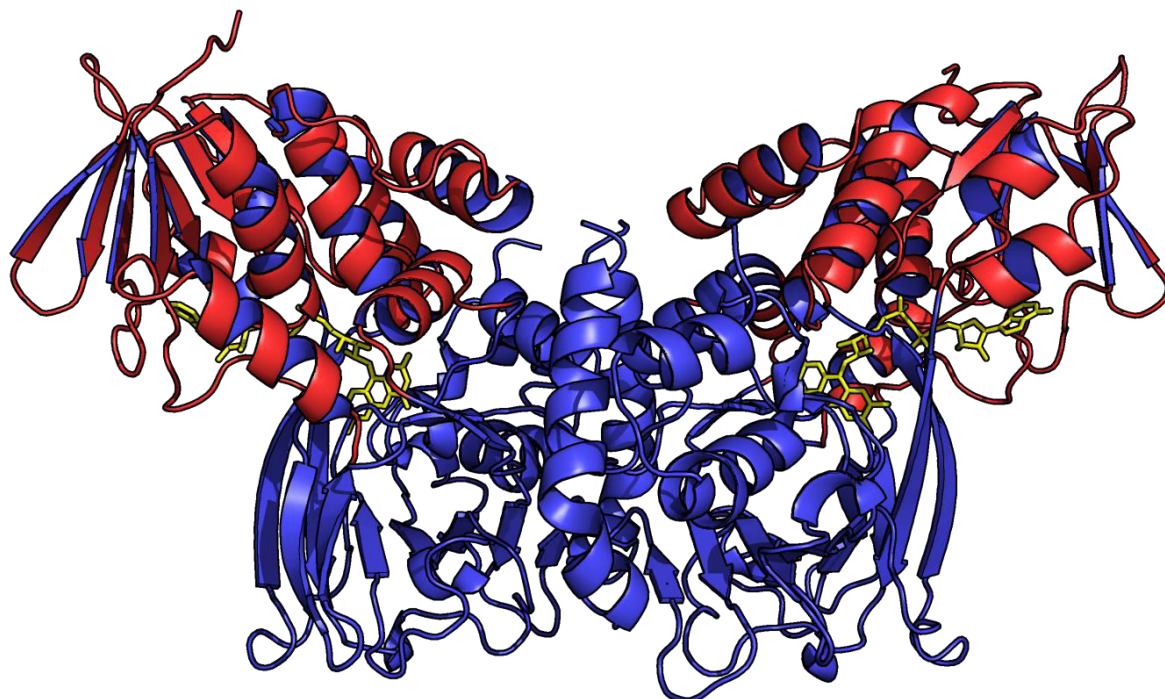


Figure 84 : Structure du mutant R326Q de la LkcE.

Le domaine de fixation du FAD est coloré en rouge, le reste de la protéine en bleu et le FAD en jaune.

Le co-facteur est stabilisé *via* un grand nombre de liaisons hydrogène impliquant notamment les résidus suivants : G17 (N), S18 (N et OG), H44 (N et ND1), E64 (N), H65 (N et O), V228 (O), H257 (NE2), T392 (N), H399 (N et ND1). Quatre molécules d'eau servent également à fixer le FAD, de même que des interactions hydrophobes avec le noyau isoalloxazine (W200, F345, V63, V284) (Figure 85).

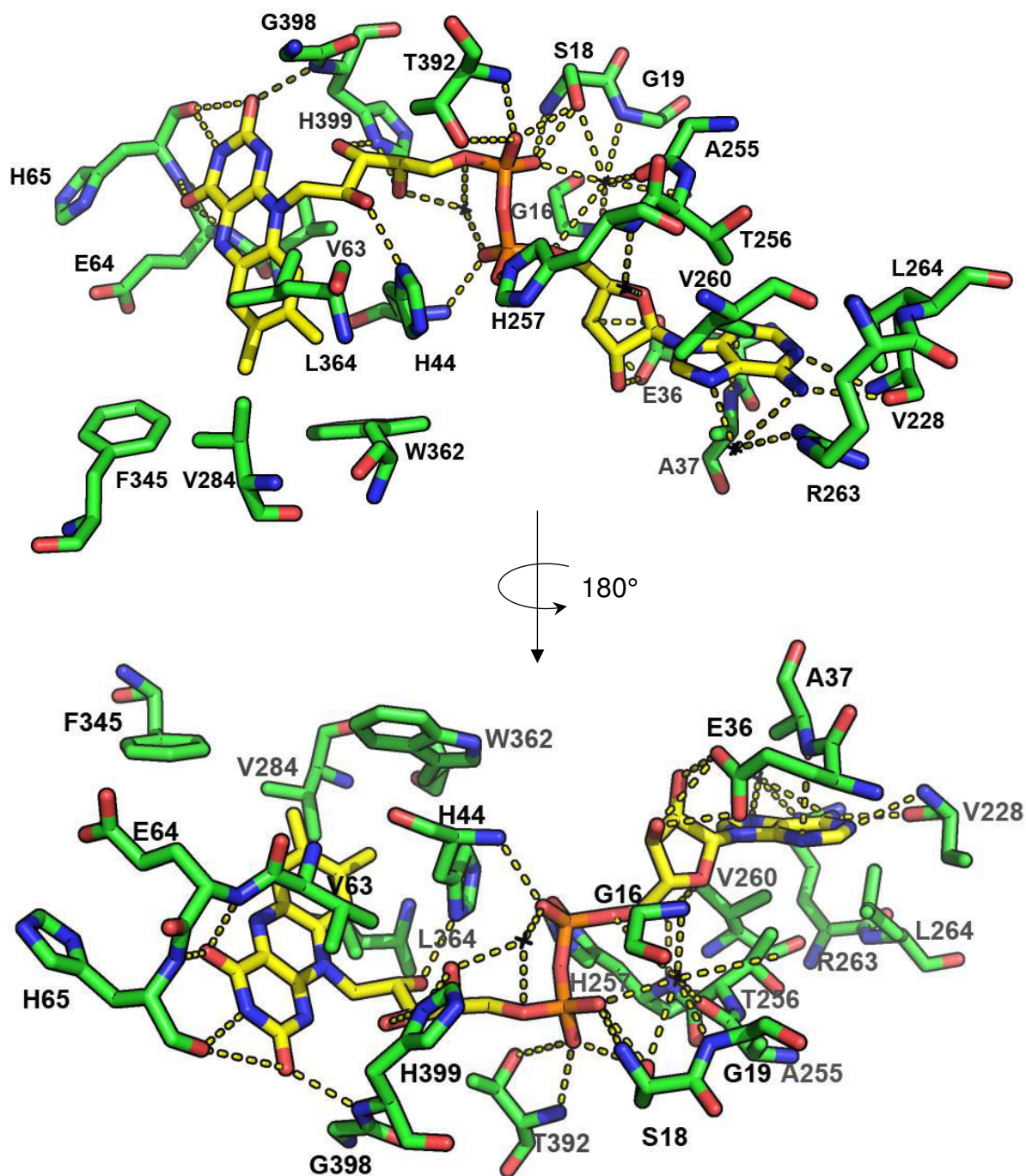


Figure 85 : Représentation du site de fixation du FAD.

Le FAD est représenté en jaune, la LkcE en vert, et les molécules d'eau par une étoile noire. Les liaisons hydrogène possibles sont représentées par un trait en pointillés jaune.

Le domaine catalytique est composé d'un long feuillet β composé de sept brins majoritairement antiparallèle (seul deux brins sont parallèle). Alors que cinq des brins sont composés de 4 à 6 résidus, deux brins sont composés de 9 et 10 résidus. Ces deux derniers brins se situent à proximité du FAD et interviennent dans la fixation du groupement isoalloxazine du FAD. Deux feuillets β supplémentaires composés de

deux brins chacun, ainsi que quatre hélices α composent également ce domaine. Le substrat de l'enzyme est également présent au site actif, bien que la densité correspondant au substrat soit légèrement moins explicite dans la chaîne B que dans la chaîne A. Néanmoins, le LC-KA05 a pu être positionné dans la densité et une *omit map* a permis de confirmer la présence du substrat. Celui-ci est sous forme linéaire dans le site actif, un état qui reflète probablement le premier stade de la catalyse (fixation du substrat dans le site actif et/ou oxydation de l'amine). A ce stade, la fonction amine et le carbone 2 du LC-KA05 sont trop éloignés pour que la réaction de cyclisation puisse s'effectuer.

Le substrat interagit avec les acides aminés du site actif à la fois par des interactions de type hydrophobe et par interactions électrostatiques. La longue chaîne aliphatique du polykétide est fixée par la Y168 avec laquelle elle interagit par stacking, la Y182, L185, le W200, la V396 et le méthyl de la T397. Il est à noter que le groupement hydroxyle de la Y182 pointe vers une partie du substrat (Figure 87).

Les interactions électrostatiques impliquent, quant à elles, la N179 qui établit une liaison hydrogène avec la fonction cétone du C3 (voir Figure 86 pour la nomenclature des atomes du LC-KA05), la Q428 qui pourrait être impliquée dans une liaison hydrogène avec la fonction acétate, et le NH peptidique de la T397 qui forme une liaison hydrogène avec la fonction ester du C1. Le groupement hydroxyle du C24 interagit avec l'E64 et par la Q326. Ceci permet de positionner la fonction amine très proche de l'azote réactif du noyau isoalloxazine du FAD pour que la réaction d'oxydation puisse avoir lieu. Deux molécules d'eau, très bien positionnées au sein du site actif, permettent également de fixer le substrat (Figure 87).

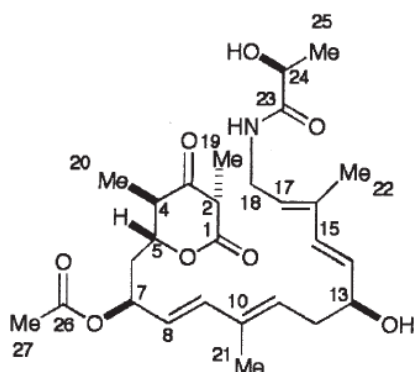


Figure 86 : Nomenclature des atomes du LC-KA05 (tiré de Arakawa et al., 2005)

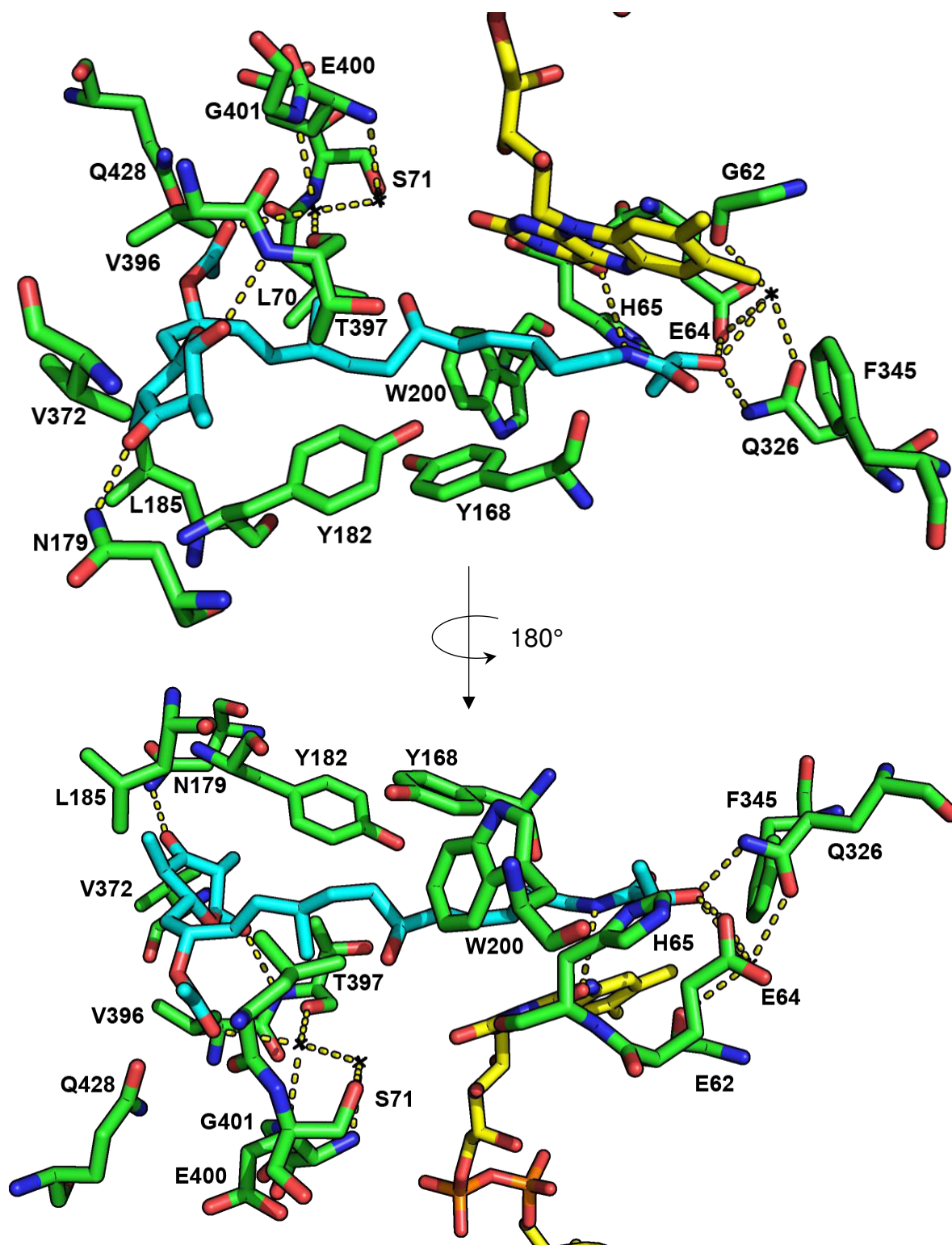


Figure 87 : Représentation du site actif du mutant R326Q en complexe avec le LC-KA05. Le substrat est représenté en cyan, l'enzyme en vert, le FAD en jaune, les molécules d'eau par des étoiles noires et les liaisons hydrogènes par des pointillés jaune. De nombreux résidus sont impliqués dans la fixation du substrat au sein du site actif.

Dans cette conformation, le site actif est relativement fermé et peu de place est disponible pour permettre la présence d'une molécule cyclique. Cependant, dans la structure de la LkcE_{wt} en P2₁ (sans molécule au sein du site actif), deux conformations pour le W200, ainsi que pour la boucle 62–71 étaient présentes, laissant penser qu'un réarrangement du site actif est possible et qui pourrait être induit par la formation de l'imine.

Deux canaux de solvant sont présents au sein de l'enzyme : l'un des canaux constitue le site de fixation du FAD (le FAD étant incorporé lors du repliement de l'enzyme), et l'autre canal sert à l'entrée et la sortie du substrat (Figure 88).

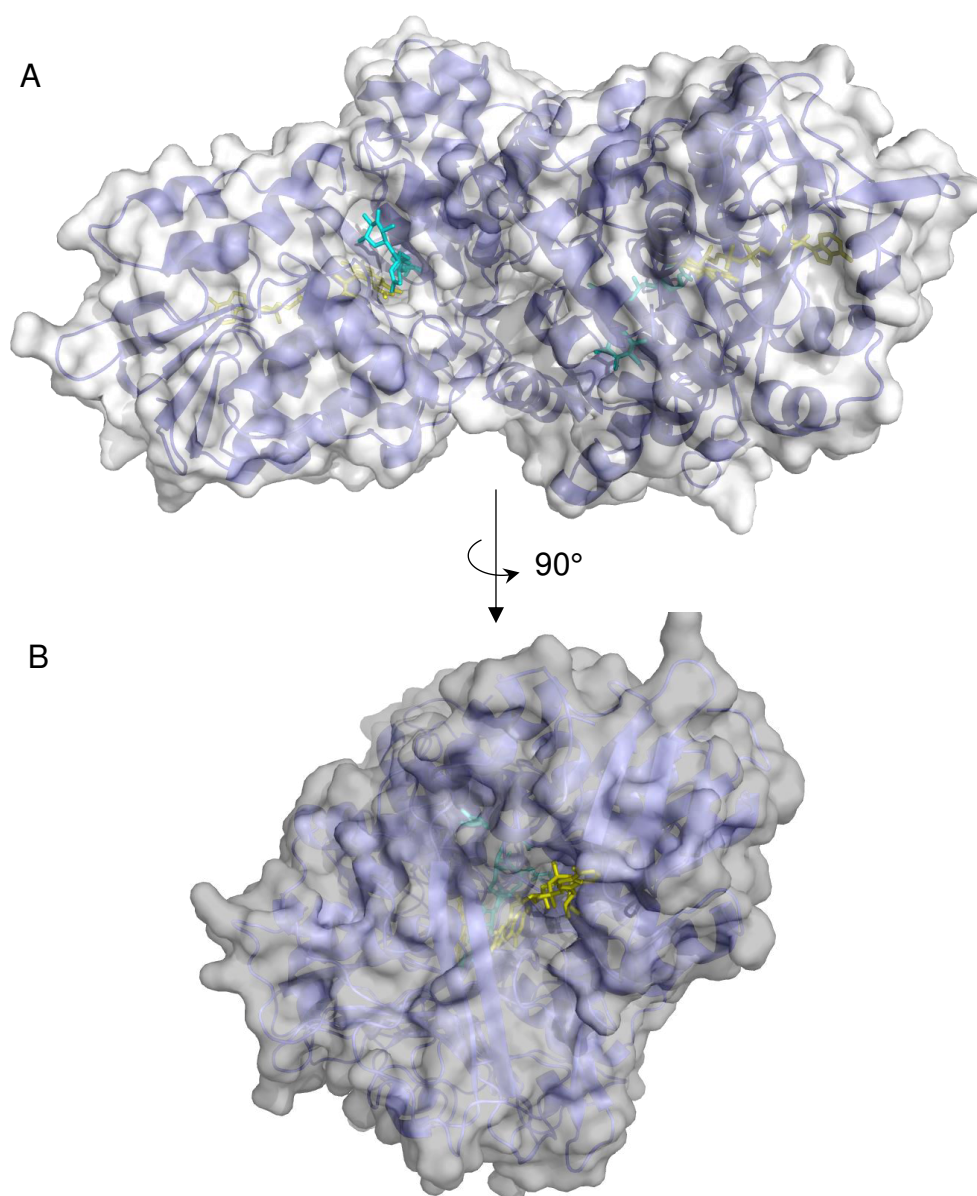


Figure 88 : Représentation en surface du mutant R326Q de la LkcE, permettant de visualiser les canaux d'entrée du substrat (A) et du FAD (B).

E) Co-cristallisation de chaque mutant et l'enzyme wt avec le LC-KA05 afin d'obtenir le produit dans le site actif

Dans le but d'obtenir la structure de l'enzyme avec le produit de la réaction au sein du site actif, l'enzyme sauvage et les deux mutants de la R326 ont été mis en présence du substrat durant 2 h à 4 °C afin de permettre une catalyse. Puis la protéine a été cristallisée par *micro-seeding*. Des cristaux ont été obtenus pour chacune des protéines, dans les mêmes conditions que les cristaux destinés au trempage avec le substrat. Les cristaux ont été congelés et plusieurs jeux de données par construction ont été enregistrés sur la ligne Proxima1.

Malgré la quantité de jeux de données enregistrés provenant de conditions de cristallisation différentes, aucun n'a permis d'identifier un éventuel produit ou même le substrat dans le site actif de l'enzyme. Seule une densité électronique affinée comme une molécule d'O₂ est présente à proximité de l'azote réactif du FAD. L'enzyme a en effet pu réagir avec le substrat et éjecter le produit de la réaction.

4. Mise au point du test d'activité de la LkcE et de la détermination des paramètres cinétiques

A) Test d'activité de la LkcE avec les analogues de substrat (DATD et EMAA)

Nous avons tout d'abord voulu connaître la capacité de l'enzyme à accepter une large variété de substrat et à catalyser des réactions variées (en l'occurrence, une condensation au lieu d'une cyclisation).

Dans un premier temps, nous avons tenté de détecter les deux analogues et l'éventuel produit formé par chromatographie en phase reverse. Un test a tout d'abord été effectué en injectant le DATD sur C18 en appliquant un gradient croissant d'acétonitrile pour l'élution. Le DATD est sorti à la fin du gradient (70% environ d'acétonitrile). L'EMAA, davantage hydrophobe que le DATD, n'a donc pas été injecté sur cette colonne. Nous avons effectué les mêmes témoins sur colonne C8. Le DATD seul est élué en un ou deux pics aux alentours de 9 minutes (Figure 89A). Cependant lors de l'injection de l'EMAA seul, nous avons constaté un pic majoritaire à 14,3

minutes et l'apparition d'un pic à 9 minutes qui n'est pas présent lors des lavages de la colonne (Figure 89B). Ceci pourrait s'expliquer par une élution de contaminants, l'EMAA n'étant pur qu'à 90%.

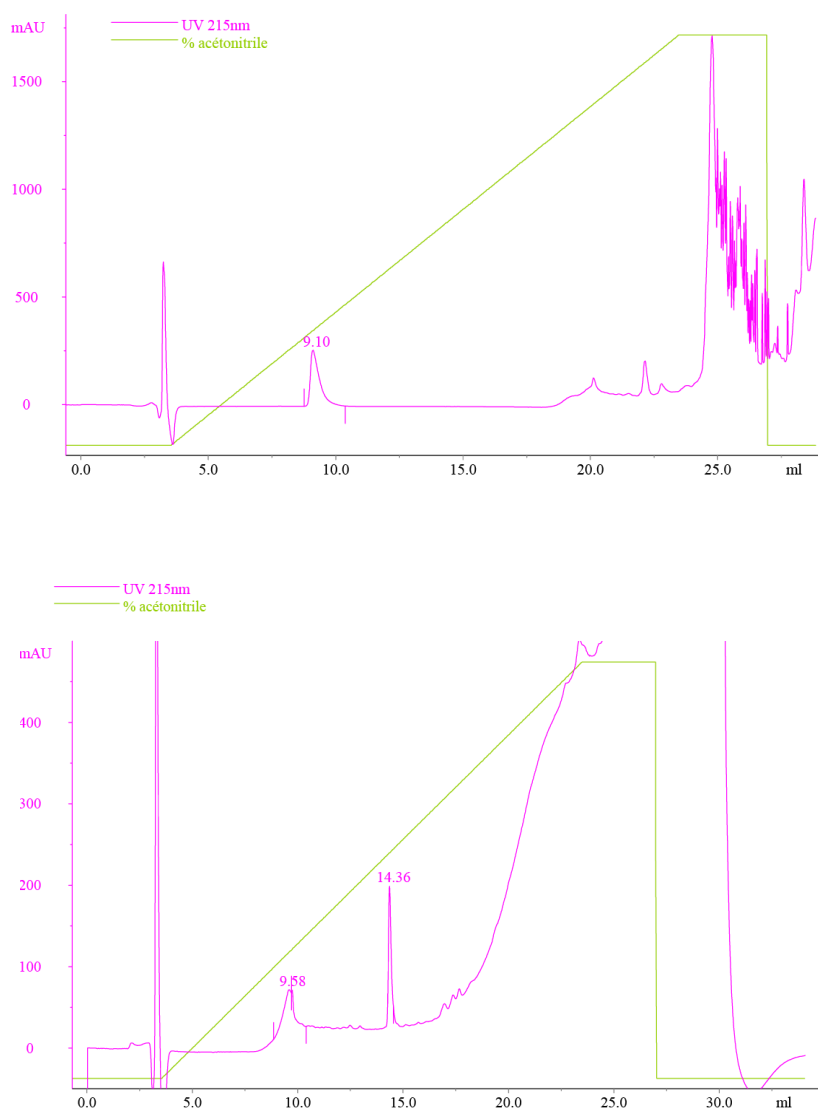


Figure 89 : Test d'élution en chromatographie en phase réverse du DATD (A) et de l'EMAA (B).

Un volume de 100 μ L de chaque analogue à une concentration de 1 mM dans le tampon de la LkcE, a été injecté sur la colonne.

Le FAD, a quant à lui, un temps de rétention de 10,5 minutes, et le TCA de 5 minutes. Chacune des molécules pouvant être séparées par C8, nous avons alors réalisé un test enzymatique en laissant 50 μ M d'enzyme réagir 1 h à 25 °C avec différentes concentration d'analogues (200 μ M, 1 mM et 50 mM). Après avoir stoppé

la réaction au TCA, et centrifugé pour faire précipiter la protéine, 100 μ L du mélange réactionnel ont été injectés. Bien que nous notions à chaque fois la présence de plusieurs pics, il est difficile de dire si une réaction a eu lieu ou non (Figure 90). On a alors laissé la réaction se prolonger sur la nuit afin de voir si les analogues sont consommés ou non. Malheureusement lors de l'injection, on retrouve seulement les pics correspondant aux analogues de substrat à une valeur d'intensité similaire à celle observée lors de l'injection des analogues seuls. Ceci laisse penser qu'aucune réaction n'a eu lieu.

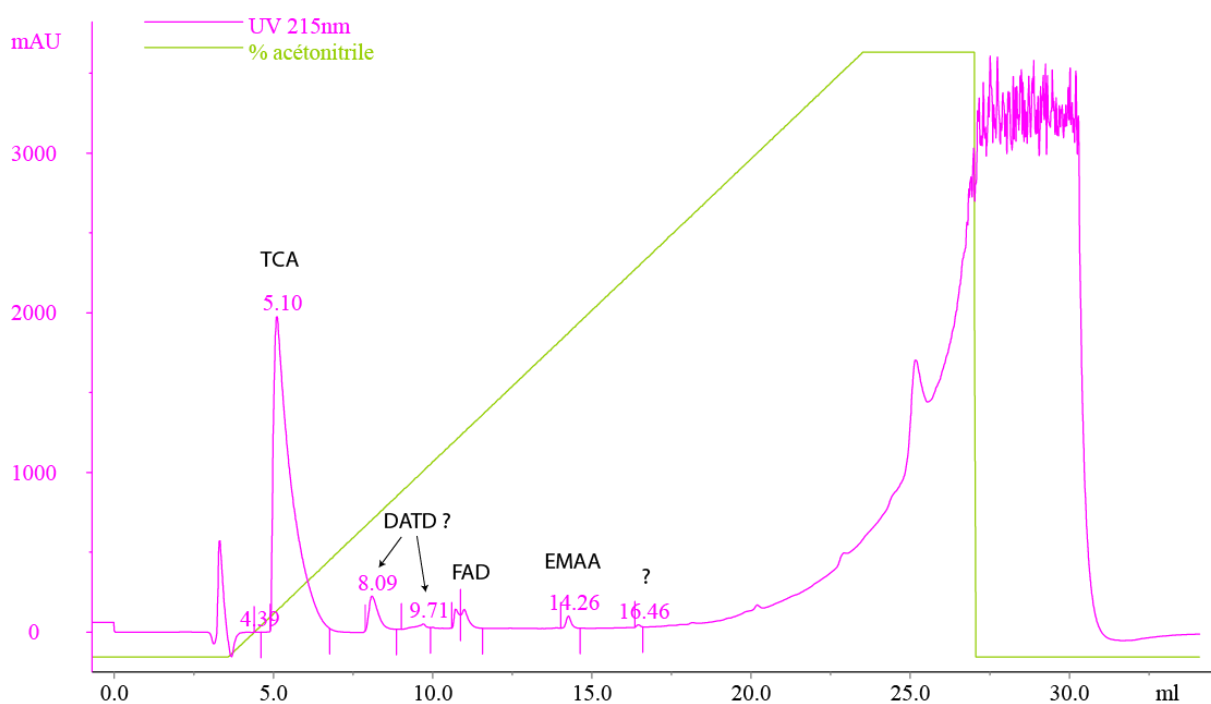


Figure 90 : Analyse par chromatographie en phase réverse du test d'activité de la LkcE (50 μ M) avec les deux analogues de substrat (1 mM chacun).

Nous avons voulu donc confirmer ou infirmer la présence de produit grâce à la LC-MS. Les tests ont été réalisés de la même façon que lors de l'analyse par chromatographie en phase réverse (50 μ M LkcE, 5 à 50 mM substrat, 2 h à 25 °C, réaction stoppée au TCA). Nous avons dans un premier temps injecté chaque analogue séparément. Alors que le DATD donne une réponse claire en masse (deux pics à 5 min et 7,5 min), l'EMAA donne une multitude de pics (Figure 91), vraisemblablement due à la présence de contaminants, ou à une élution en plusieurs pics de l'EMAA (un seul pic est visible en UV).

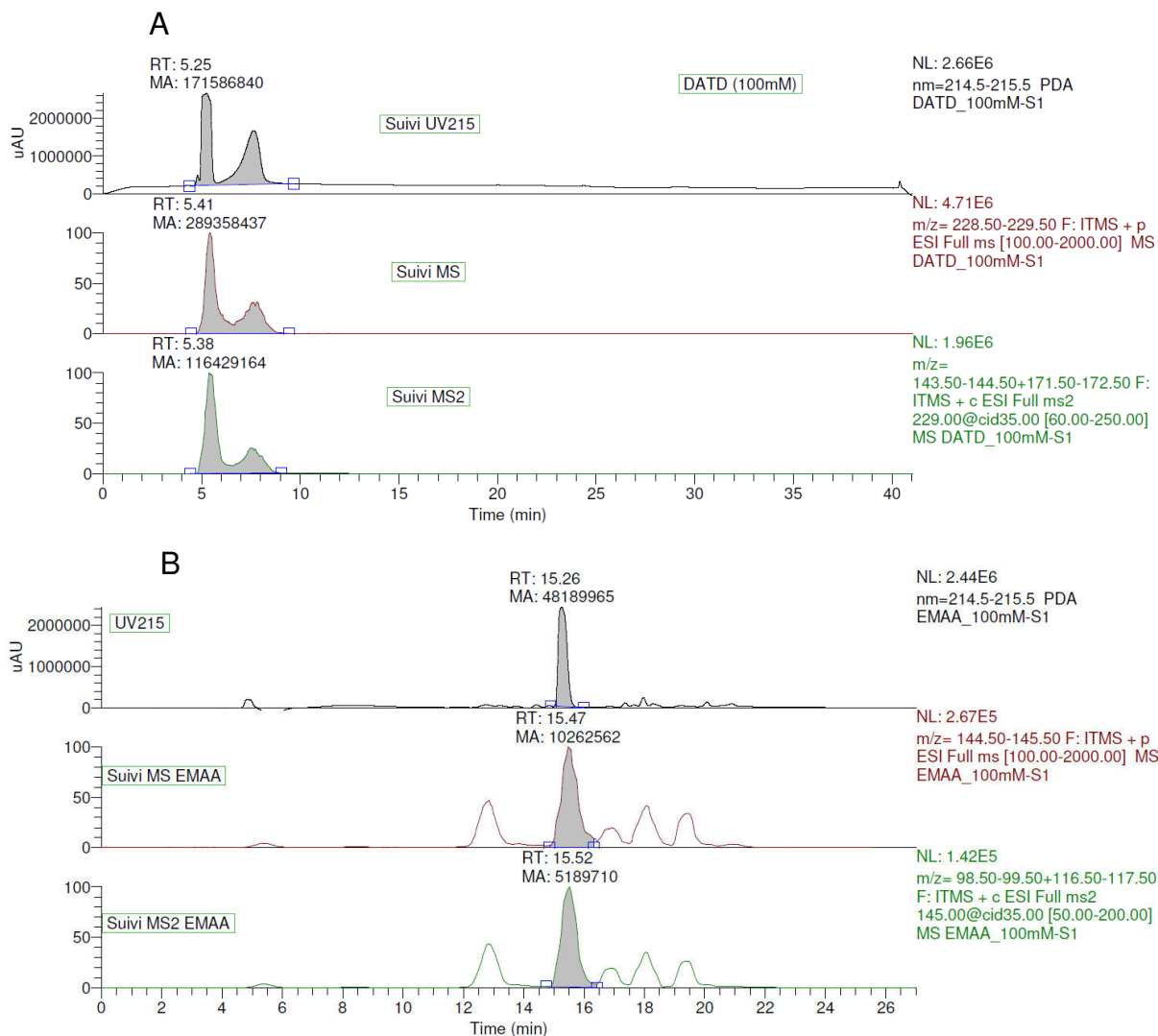


Figure 91 : Suivi en UV, MS et MS² des deux analogues de substrat : en A) le DATD et en B) l'EMAA.

Nous avons tenté de suivre le produit en masse, puisque la masse du composé attendu est de 370 Da. Un pic est visible à un temps de rétention (RT) de 21,98 min (Figure 92). Cependant, ce pic est de faible intensité et les pics UV correspondant à l'EMAA et au DATD demeurent toujours identifiables (Figure 92A). De plus, des spectres MS2 et MS3 mettent en évidence des ions fils de masse inattendue.

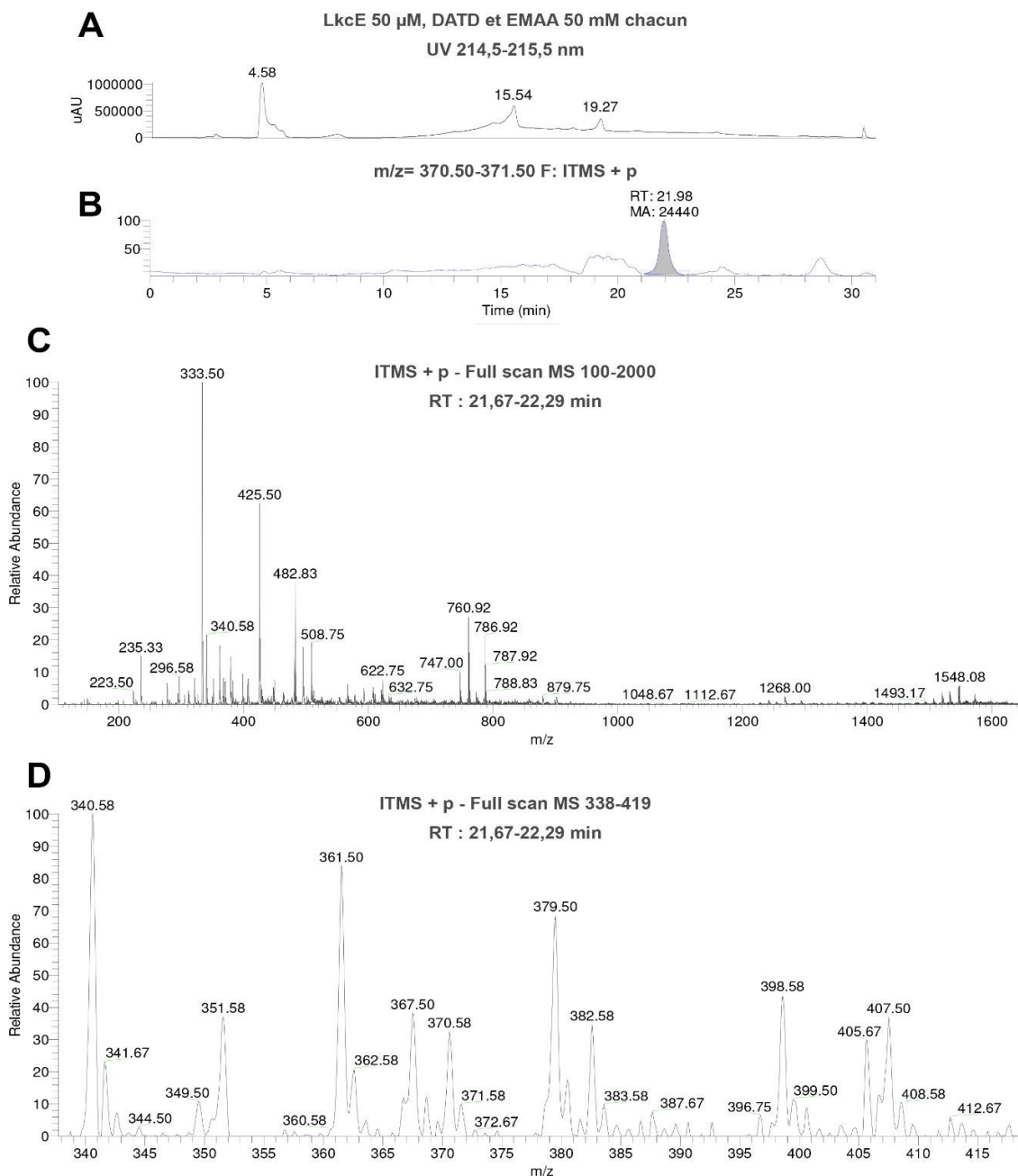


Figure 92 : Suivi par spectrométrie de masse du produit de la réaction attendu.

A : Suivi UV à 215 nm du produit de la réaction. Le pic à 4,58 min pourrait correspondre au DATD et le pic à 15,58 min à l'EMAA. Un pic de faible intensité est visible à 19,27 min dont la masse ne correspond pas au produit attendu.

B : Suivi LC-MS du produit de la réaction. Le produit de la réaction est suivi en mode positif sur une gamme de m/z de 370,50 à 371,50 qui comprend la valeur de m/z attendue pour le produit (371 [M+H]⁺) Un pic apparaît à 21,98 min, bien qu'aucun pic ne corresponde en UV.

C : Scan par spectrométrie de masse sur le pic identifié en LC-MS (RT : 21,67 à 22,29 min). Aucun pic de valeur de $m/z = 371$ ne peut être clairement identifié.

D : Zoom sur le scan par spectrométrie de masse précédent ($m/z = 338-419$). Deux pics de valeur de $m/z = 370,58$ et $371,58$ sont présents. Cependant, leur intensité apparaît extrêmement faible ce qui rend compte de la faible proportion de ce composé au sein du pic.

Nous avons alors examiné la présence de ce pic dans trois témoins négatifs : la LkcE avec le DATD seul, l'enzyme avec l'EMAA seul et d'un mélange composé de DATD, d'EMAA et de la LkcE précédemment chauffée à 100 °C afin de dénaturer l'enzyme. Nous avons retrouvé la présence de ce pic de $m/z = 371$ dans l'EMAA et dans le témoin négatif de la réaction (Figure 93).

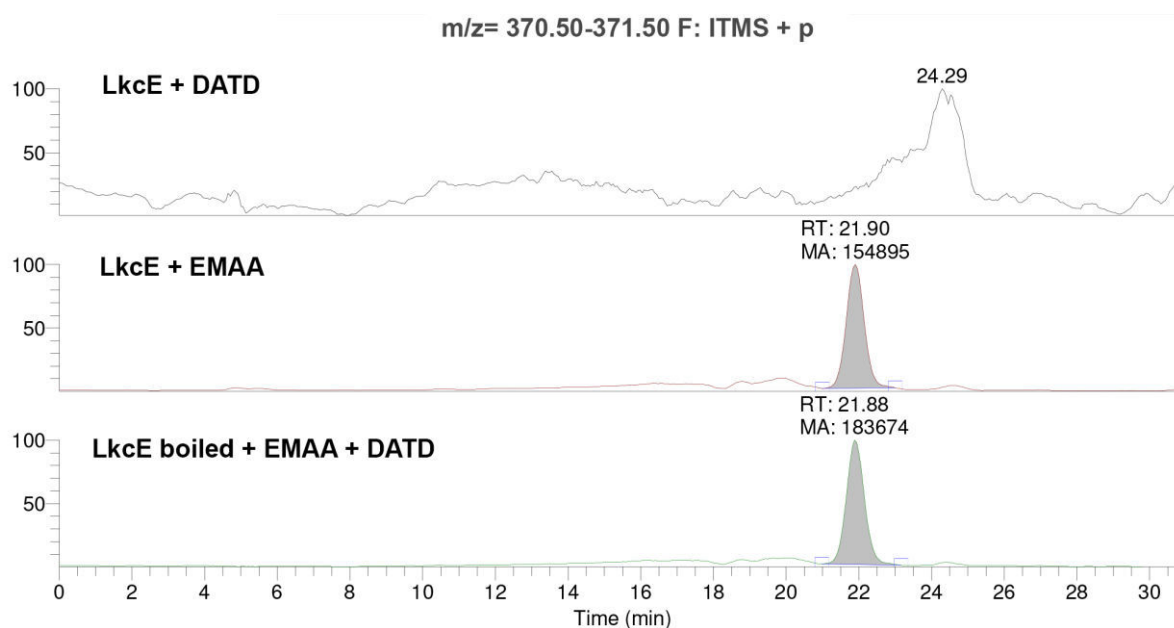


Figure 93 : Analyse LC-MS des témoins négatifs.

Un suivi des composés de $m/z = 370$ a été réalisé pour chaque échantillon. Nous constatons que pour l'EMAA, que ce soit en absence de DATD ou en absence d'enzyme, un pic correspondant à la masse attendue est présent, ce qui indique qu'il ne s'agit pas du produit de la réaction mais un contaminant présent dans l'EMAA.

B) Test d'activité de la LkcE avec le substrat naturel

Devant l'impossibilité de détecter un produit formé à partir des deux analogues de substrat, nous avons entrepris de mettre au point un test permettant de mesurer l'activité de la LkcE avec son substrat naturel – le LC-KA05 dorénavant purifié par les chimistes allemands. En vue de la faible quantité de substrat à notre disposition, il était nécessaire d'avoir recours à une méthode présentant une forte sensibilité. Nous avons donc choisi la LC-MS comme méthode d'étude. La mise au point de cette méthode pour l'étude de l'activité de la LkcE avait également débuté grâce aux analyses faites sur les analogues de substrat, ce qui représentait également un avantage conséquent.

Le but était de suivre la disparition du substrat par LC-MS en quantifiant cette consommation et de suivre la formation du produit. Tout d'abord, nous avons cherché à quantifier le substrat afin de pouvoir réaliser une gamme étalon. Pour cela, il a fallu se placer à l'identique des conditions qui ont ensuite été utilisées pour les tests cinétiques. Le substrat a donc été placé dans le tampon de la protéine (30 mM HEPES, 150 mM NaCl, 1 mM EDTA (pH 7,5)), et immédiatement extrait volume à volume dans l'acétate d'éthyle. Deux extractions ont été effectuées sur la phase aqueuse. Une fois le solvant évaporé, puis repris dans un mélange eau:acétonitrile 80:20, les échantillons ont été analysés par LC-MS. Une gamme étalon a ainsi pu être établie, et une équation représentant la courbe obtenue a été déterminée (Figure 94).

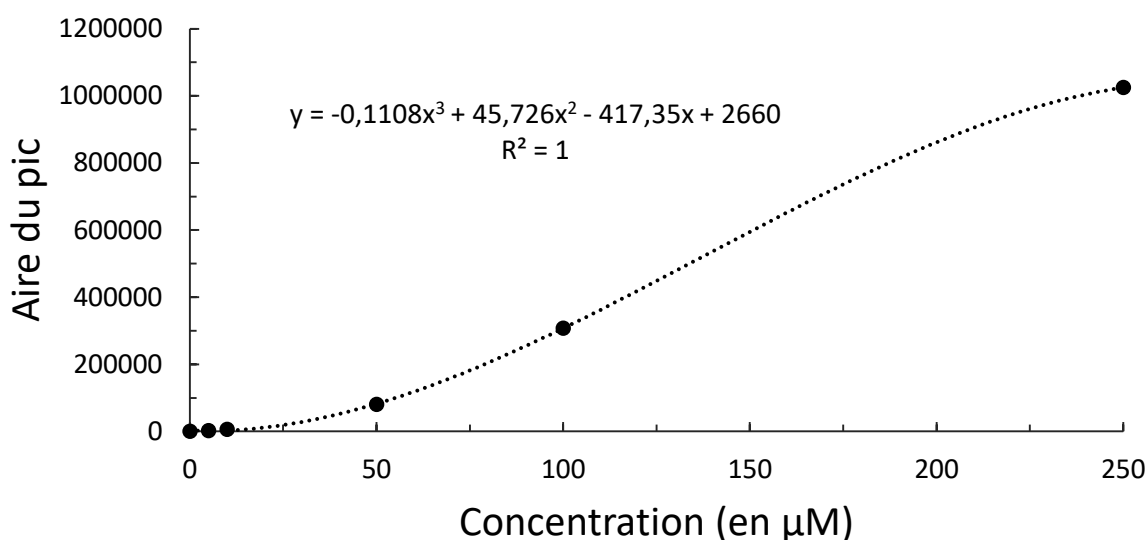


Figure 94 : Courbe étalon reliant l'aire du pic correspondant au LC-KA05 ($m/z = 504,5$ en mode négatif) à la concentration en substrat (en μM).

La courbe (non linéaire sur les concentrations utilisées) est ajustée à un polynôme du 3^{ème} degré

Afin de déterminer les paramètres cinétiques de la LkcE, l'enzyme (100 nM) a été mise en présence de 250 μM de substrat, et la réaction a été stoppée à différents temps en prélevant une fraction aliquote et en la diluant dans l'acétate d'éthyle afin de précipiter la LkcE. La phase aqueuse a été extraite une seconde fois. Une fois le substrat évaporé, les échantillons correspondant à différents temps ont été analysés et quantifiés par LC-MS (Figure 95).

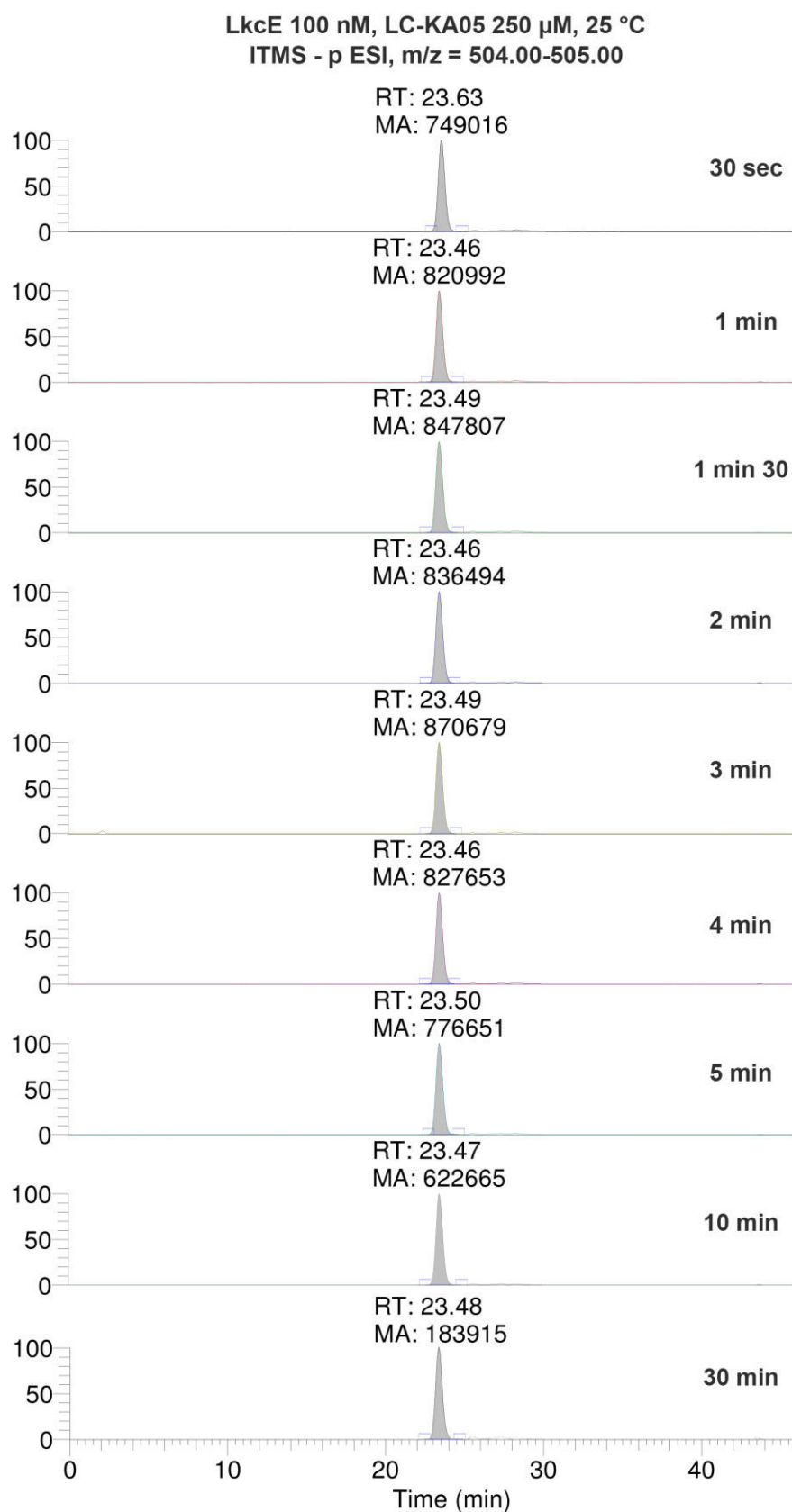


Figure 95 : Suivi par LC-MS de la consommation du substrat par la LkcE.

La réaction a été suivie de 30 s à 30 min et l'aire du pic correspondant au LC-KA05. Nous constatons une diminution de l'aire du pic (MA) à partir de 3 minutes.

Cependant, du fait de l'hydrolyse du substrat pendant les tests d'activité (voir partie III.3.D)a), nous avons également suivi la cinétique d'apparition de la forme hydrolysée du substrat ($m/z = 462,5$). En effet, nous ne savions pas si l'enzyme était capable de réagir avec cette forme du substrat. Les analyses par LC-MS suggère que dans ces conditions expérimentales, l'hydrolyse semble lente et la quantité de forme hydrolysée est négligeable devant le substrat intact (Figure 96).

Le suivi de la disparition du substrat nous a permis déterminer un k_{obs} , c'est-à-dire la constante de vitesse de la réaction, dans ces conditions de concentration, et dans les conditions de vitesse initiale. Nous avons utilisé une concentration en substrat présumée saturante pour l'enzyme afin de pouvoir estimer une V_{max} puis en déduire une estimation du k_{cat} ($k_{cat} = V_{max}/[E]_0$) de l'enzyme. Pour cela nous avons mis en présence 100 nM de LkcE avec 250 μ M de substrat. Cette concentration permet en plus de rester dans la zone linéaire de quantification en LC-MS. Pour chaque mesure, nous avons converti l'aire du pic de masse en concentration de LC-KA05 grâce à la courbe étalon. Cependant, on constate une phase de latence d'environ 3 minutes durant laquelle la concentration en substrat ne diminue pas. En excluant cette phase de latence et on ne conservant que la partie linéaire de la courbe (qui représente donc la période où l'enzyme est en condition de vitesse initiale), nous avons pu ajuster la courbe à une régression linéaire. Cela nous a permis de calculer une vitesse initiale V_i de 6 μ M.min⁻¹. En divisant par la concentration en enzyme, nous avons déterminé un k_{obs} de 1 s⁻¹ (Figure 97). En supposant que l'enzyme fonctionne à saturation, ce k_{obs} reflète le k_{cat} de la réaction. Cette valeur de k_{obs} suggère que cette étape ne serait pas limitante dans la voie de synthèse du polykétide puisque la constante cinétique de l'ensemble de la PKS serait de l'ordre de la min⁻¹.

En revanche, il était impossible de mettre en évidence la formation du produit de la réaction, que ce soit en mode positif ou négatif. Nous avons alors tenté de suivre le produit qui pourrait être formé si le groupement acétate était hydrolysé de la même façon que pour le substrat, sans davantage de réussite. Un suivi des ions fils potentiels dus à la fragmentation par l'ionisation ou des différents adduits s'est également révélé infructueux.

A ce stade, devant la faible fiabilité de la méthode, nous avons préféré suspendre les analyses par LC-MS, afin de réfléchir à un test plus pertinent et moins couteux pour déterminer plus rigoureusement les paramètres cinétiques de l'enzyme.

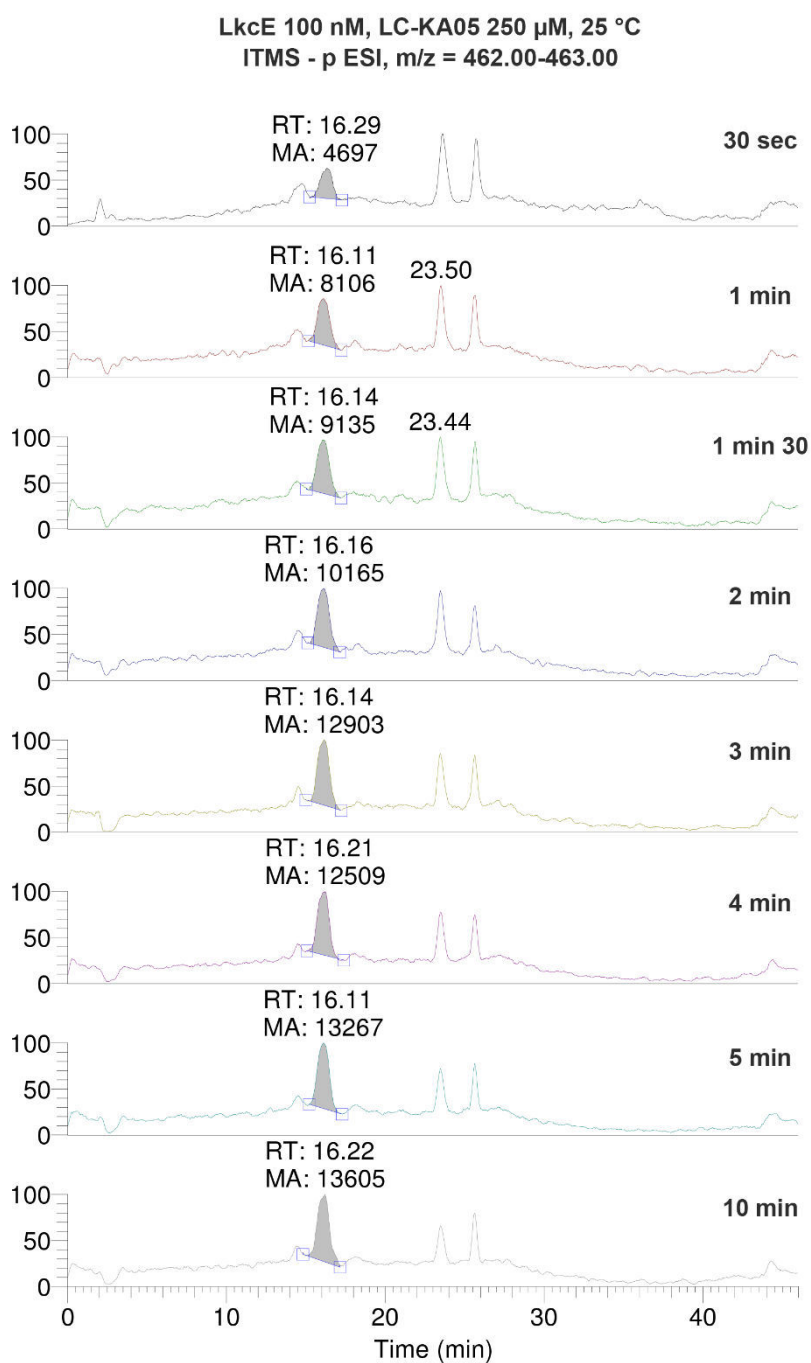


Figure 96 : Suivi par LC-MS de l'apparition de la forme hydrolysée du substrat.

La formation de ce produit indésirable a été suivie de 30 s à 30 min (le temps t = 30 min n'est pas affiché ici (RT : 16,16 ; MA : 6745) car les données MS sont très bruitées).

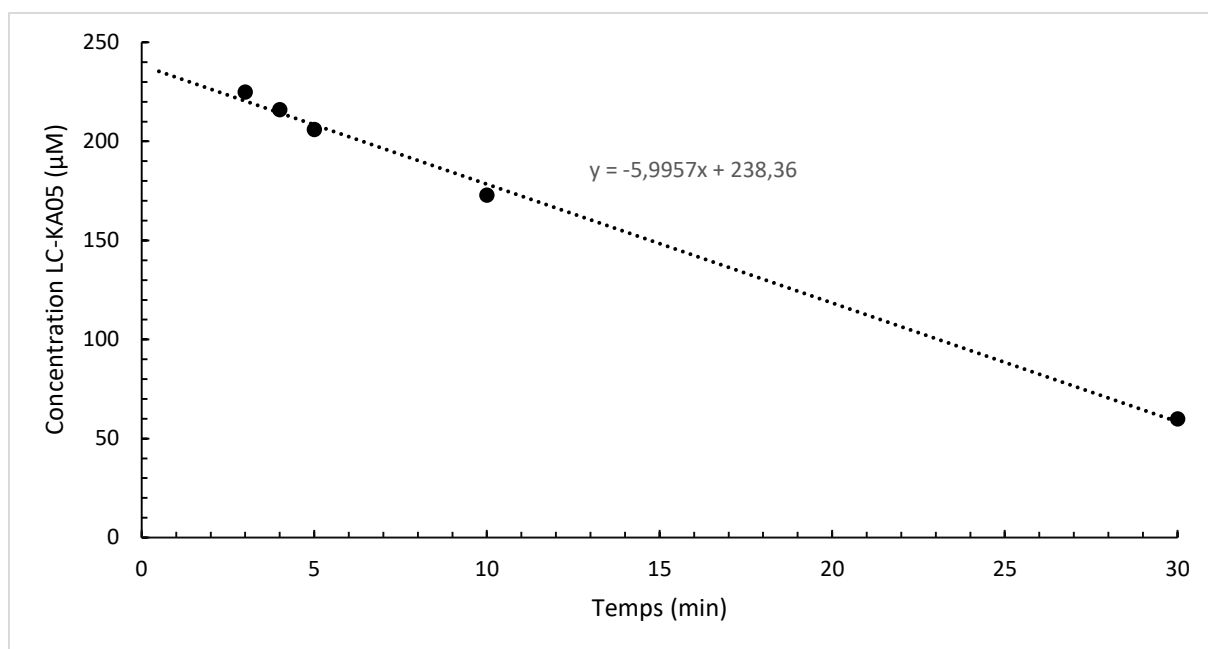


Figure 97 : Courbe représentant l'intensité obtenue par LC-MS en fonction du temps de la réaction.

Les données ont été ajustées à l'équation d'une droite en prenant les temps de 3 min à 30 min.

*DISCUSSION ET
PERSPECTIVES*

Structure à faible résolution d'un module de PKS *trans*-AT : une conformation ouverte inattendue

Jusqu'à présent, aucune structure à haute résolution n'a pu être obtenue pour un module de PKS modulaire. Les méthodes classiques permettant de résoudre des structures à haute résolution (RMN et cristallographie aux rayons X) ne sont pas adaptées à l'étude de ces systèmes complexes, de haut poids moléculaires (RMN) et dotés d'une grande flexibilité intrinsèque. Dès lors, l'utilisation de méthodes complémentaires, permettant la résolution de structure à plus faible résolution, s'impose pour l'étude des PKS. Le SAXS constitue une technique particulièrement bien adaptée à l'étude d'objets flexibles. Elle permet de déterminer l'enveloppe moléculaire de la protéine, c'est-à-dire sa conformation globale, puis de placer des structures à haute résolution au sein des enveloppes obtenues, renseignant sur l'organisation de protéines à plusieurs domaines et de protéines complexes. Notamment, elle permet également d'étudier la dynamique des protéines en solution.

Nous avons utilisé une approche visant à étudier dans son intégralité le module 5 de la PKS synthétisant la virginiamycine M, mais aussi des sous-ensembles de ce module. En étudiant la structure et la dynamique de portions individuelles du module, nous avons été capables de reconstruire l'architecture et le comportement de l'ensemble. Ces détails n'auraient pas été accessibles juste à partir de l'analyse du module entier.

Le module 5 affiche une conformation « bras ouvert » dans laquelle les deux ACP occupent une position non équivalente. L'ACP_{5a} est situé directement à proximité du KS₅, semble participer directement à la condensation. L'ACP_{5b}, quant à lui, est localisé à distance de l'ACP_{5a} à l'extrémité d'un *linker* flexible, et est donc plus accessible que l'ACP_{5a}. Il devrait donc jouer un rôle différent, et intervenir en série après l'ACP_{5a} (Rahman, *et al.*, 2005). L'ACP_{5b} pourrait alors servir de support pour les réactions de modification impliquant la cassette de β -méthylation et/ou pour le transfert du polykétide vers le module suivant. Afin de valider cette hypothèse, il est nécessaire de muter tour à tour chaque ACP et de vérifier si (1) la condensation reste possible, (2) si la β -méthylation est toujours effectuée et (3) si la chaîne en élongation est toujours transférée vers le module en aval. Pour cela, nous avons déjà procédé à des

mutagenèses dirigées pour muter la sérine canonique des ACP, puis les mutants du module 5 ont été produits et purifiés.

Deux aspects de la caractérisation ont été abordés. Dans un premier temps, nous avons tenté de modifier le module 5 et chacun des deux mutants soit au niveau des ACP soit au niveau du domaine KS₅. Les ACP ont en principe été modifiés par ajout d'un bras phosphopantéthéine portant une unité malonate. Il sera néanmoins critique de vérifier si les modifications ont été effectives ou non par protéolyse ménagée suivie de la spectrométrie de masse. Cette modification a été effectuée en utilisant du malonyl-CoA, transféré sur les ACP par la Sfp, une PPTase de spécificité large (Quadri, *et al.*, 1998). Le KS₅ a, quant à lui, été mis en présence du crotonyl-SNAC (S-N-acétylcystéamine), ce qui va permettre le transfert de crotonate sur la cystéine catalytique. Ceci devrait mimer la présence de la chaîne en extension au sein du site actif du KS₅. Il est à noter que lorsque le KS₅ est acylé, nous avons modifié les ACP par ajout du groupement phosphopantéthéine à partir de CoA, pour obtenir les ACP sous leur forme *holo* (la réaction tournerait si une unité malonate était présente sur l'ACP). Nous avons enregistré des jeux de données SAXS sur ces constructions dans le but d'isoler des conformères différents de celui obtenu pour le module 5 sous sa forme *apo*. Cela permettrait d'obtenir des « clichés » des différents états conformationnels du module 5 au cours de la catalyse. Ces données SAXS sont actuellement en cours de traitement.

Un deuxième aspect consistera à modifier le KS₅ et les deux ACP de la même façon que précédemment, sauf que les modifications du KS₅ et des ACP se feront de façon simultanée afin que la catalyse puisse avoir lieu. Le produit de condensation sera ensuite libéré par hydrolyse alcaline, extrait à l'acétate d'éthyle, puis modifié chimiquement par ajout de TMS-diazométhane pour méthyler le produit afin de le rendre plus volatile pour être analysé en GC-MS, en utilisant un standard préparé par le laboratoire du Pr. Kirschning. La finalité de travail sera d'identifier si un produit est formé par le mutant de l'ACP_{5a} et celui de l'ACP_{5b} et éventuellement de quantifier ce produit. Cela permettra d'identifier si les deux ACP sont impliqués dans la condensation ou s'ils ont effectivement des rôles non équivalents.

Un autre aspect intéressant consiste en l'organisation des *linkers* du module 5. En effet, le long *linker* de 298 acides aminés reliant le KS₅ à l'ACP_{5a} semble ordonné. Ce

linker est composé d'une portion homologue au *linker* reliant le KS à l'AT chez les PKS de type *cis*, d'une portion homologue au *linker* post-AT (Tang, *et al.*, 2006 ; Tang, *et al.*, 2007) et de 100 résidus qui ne correspondent à aucune région de structure connue mais qui semblent être structurés. Bien que l'organisation globale du module 5 soit similaire à celle obtenue par cryo-EM sur la PKS de la pikromycine (Dutta, *et al.*, 2014), en respectant cette forme en arche, le linker post KS diffère dans son orientation. Malgré la rotation des *linkers* effectuées grâce au programme SASREF, la position de ces derniers par rapport au KS₅ diffère de celle visible dans la structure cryo-EM de PikAIII. Pour expliquer une telle différence, il a été proposé que la position de ce *linker* serait sous la dépendance de la présence ou de l'absence d'un domaine de dimérisation en C-terminale (Dutta, *et al.*, 2014). Ceci expliquerait pourquoi ce *linker* se situerait sur le côté du domaine KS dans les structures cristallographique des didomaines KS-AT sortis de leur contexte modulaire. Cependant, plusieurs PKS de type *trans*-AT (y compris la virginiamycine) ne disposerait pas de tels motifs de dimérisation C-terminale, ce qui expliquerait l'agencement du *linker* et donc la conformation ouverte du module 5 de la PKS synthétisant la virginiamycine M.

Le *linker* reliant les deux ACPs, bien que plus court (seulement 28 acides aminés) semble quant à lui très flexible et de conformation étendue. Ceci peut s'expliquer par le besoin de rendre l'ACP_{5b} plus accessible pour l'interaction avec les partenaires en *trans* vraisemblablement, mais aussi pour permettre à cet ACP d'interagir avec le reste du module 5 (ACP_{5a} et/ou KS₅). La conformation ouverte du module 5 serait ainsi favorable pour l'interaction en *trans* avec les différents partenaires du module : que ce soit la cassette de β -méthylation (sous forme de complexe ou bien alors les protéines de la cassette agissant tour à tour) ou bien encore avec le module 6.

Un autre point important qu'il reste à déterminer est la structure des modules arborant un panel de domaines de modification optionnels (DH qui est dimérique au moins chez les systèmes *cis*-AT et ER notamment (monomérique chez les *cis*-AT PKS)). Pour cela, nous avons entrepris la caractérisation par une approche couplée SAXS et cryo-EM de plusieurs modules de la PKS synthétisant la virginiamycine M. Ainsi le module 1 (KS-DH-KR-MT-ACP-ACP), le module 2 (KS-DH-KR-ACP), le module 4 (KS-DH-KR-ACP), ainsi que VirFG entier (Figure 98) ont été clonés par Sabrina Collin et le Dr. Christophe Jacob. Les protéines ont été purifiées par Sabrina Collin et moi-même. Les purifications de ces protéines sont laborieuses car les

modules, en plus d'être peu solubles lors de la lyse cellulaire, sont composés de nombreux *linker* et sont de ce fait sujet à une protéolyse importante. L'utilisation de pastille anti-protéase (cOmplete™ ULTRA Tablets, Mini, EDTA-free, EASYpack Protease Inhibitor Cocktail, Roche), l'ajout de fluorure de phénylméthylsulfonyl (PMSF), et d'EDTA suite à l'IMAC s'avère nécessaire pour limiter la protéolyse. Des données SAXS sur ces protéines ont été enregistrées et doivent encore être traitées. Le module 2 et le module 5 ont également été envoyées chez nos collaborateurs (plateforme de Microscopie Electronique, AFMB (UMR7257), Marseille, gérée par le Dr. Silvia Spinelli et le Pr. Christian Cambillau). Les finalités de ce sujet sont multiples : tout d'abord, identifier le positionnement des domaines de modification optionnels ainsi que leur influence sur la structuration globale des modules. Ensuite, l'étude de VirFG entier permettrait de caractériser pour la première fois un di-module d'une *trans*-AT PKS (module 6–module 7) au sein d'une même protéine et ainsi de mieux comprendre le transfert du polykétide d'un module à l'autre. Et enfin, la caractérisation systématique et comparative de toutes les protéines permettrait de proposer un modèle d'organisation globale de la PKS, surtout au niveau des modules incorporant des panels différents de domaines.

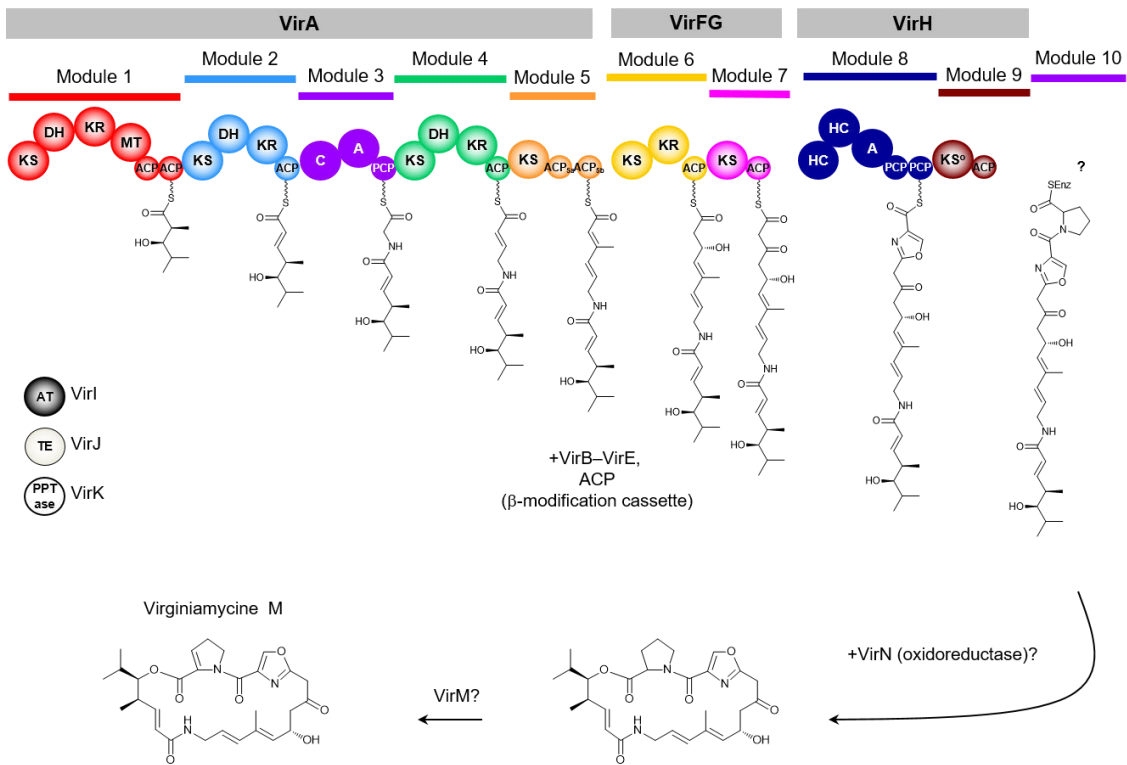


Figure 98 : Rappel de l'organisation en module et domaine de la PKS synthétisant la virginiamycine M.

Une sous-unité provenant de la PKS synthétisant la lankacidine C est également à l'étude. En effet, la LkcA constitue en elle-même une hybride PKS-NRPS et représente un excellent modèle d'étude pour comprendre le mécanisme d'interaction entre les modules NRPS et les modules PKS. De plus, les modules NRPS se structurent sous la forme de monomères (Sieber, *et al.*, 2002) et il serait intéressant de suivre l'état oligomérique de ces modules dans un contexte dimérique. Des premières données SAXS et cryo-EM sont déjà disponibles pour cette construction, dans laquelle la partie NRPS de la protéine semble être monomérique alors que la LkcA s'homodimérise par l'intermédiaire du KS.

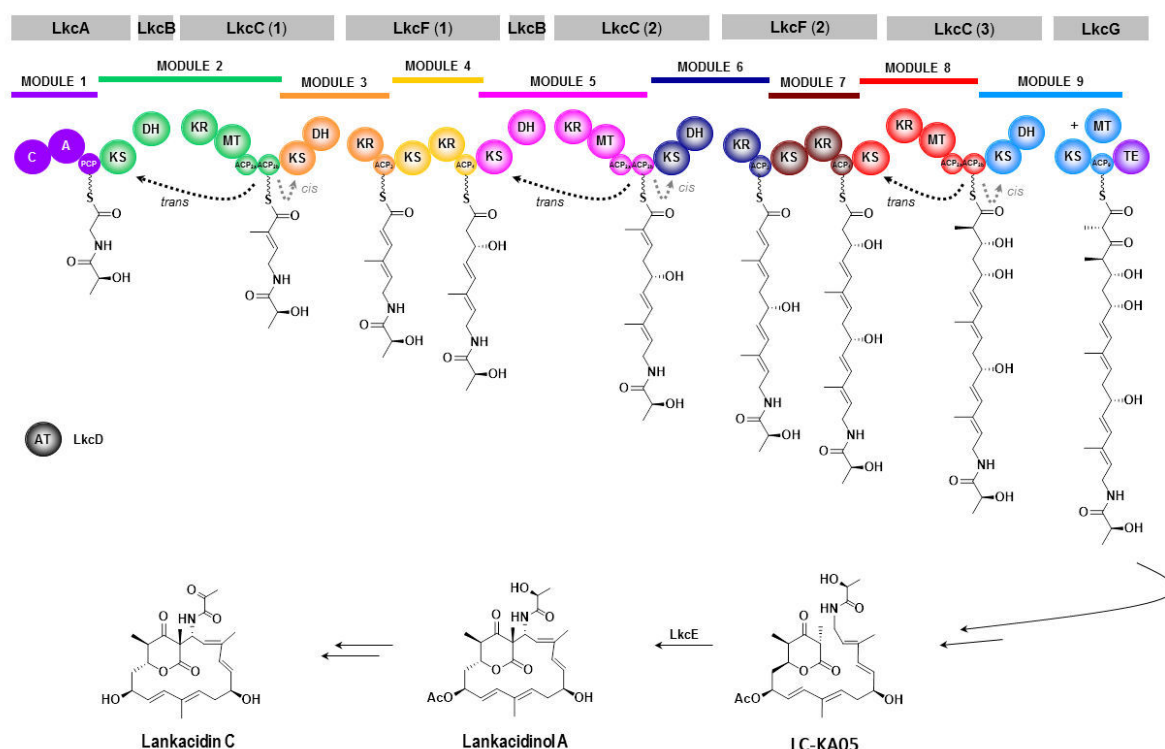


Figure 99 : Rappel de l'organisation de la PKS synthétisant la lankacidine C.

La LkcA présente l'avantage de constituer une protéine discrète ; ainsi les bornes n'ont pas besoin d'être délimitées, ce qui augmente les chances d'avoir la protéine soluble. D'autre part, le module NRPS se trouve être le premier module de la PKS. Sa modification par son substrat naturel en sera donc grandement facilitée, puisqu'il s'agit de la glycine.

Une autre cible d'études potentielle est la sous-unité Bam 5924 de la PKS de type *trans*-AT synthétisant l'ényacyloxine (un antibiotique de la classe des polyènes) (Figure 100). Cette sous-unité comprenant 2 modules présente l'avantage d'avoir un domaine KS en N-terminale et en C-terminale. Ainsi il y aurait deux domaines dimériques de chaque côté de la sous-unité, ce qui pourrait rendre plus rigide la protéine et donc

limiter sa sensibilité à la protéolyse. De plus, cela permettrait d'étudier l'organisation de deux modules entre eux, sur une protéine de taille moléculaire moindre que VirFG. Le module 3 de cette PKS dont le KS fait partie de Bam 5924 forme un module scindé sur deux chaînes polypeptidiques et l'étude de l'interface de docking entre ces deux sous-unités est particulièrement intéressante. La caractérisation de cette interface de docking fait partie du sujet de thèse de Fanny Risser au sein de notre laboratoire.

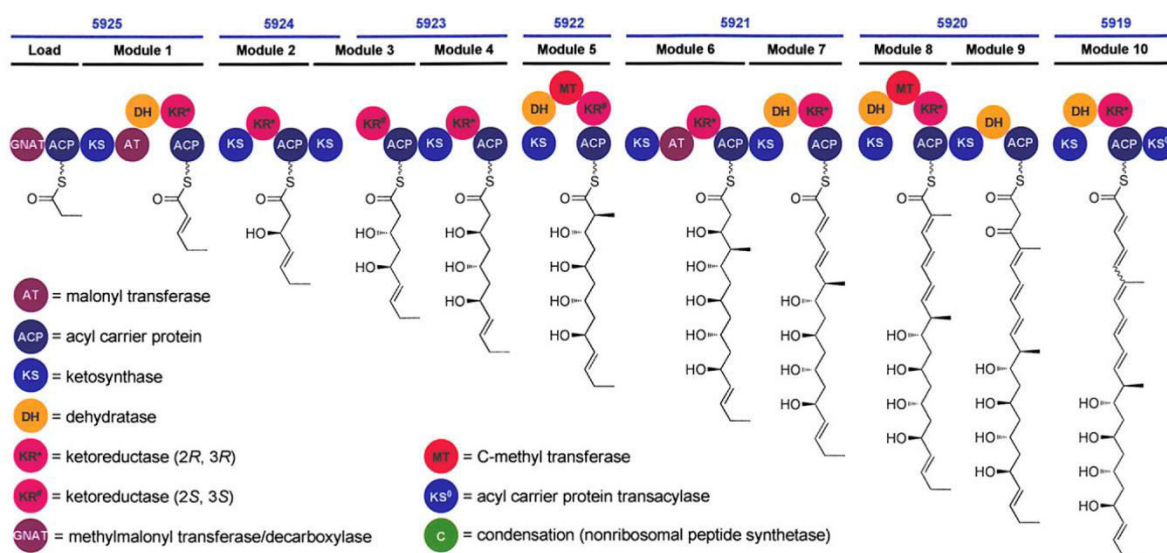


Figure 100 : Organisation de la PKS synthétisant l'énacyloxoine (Mahenthiralingam, et al., 2011).

Cette PKS s'organise en 7 chaînes peptidiques et 10 modules. Le module 3 est à cheval sur les sous-unités 5934 et 5923. Au vu des analyses de séquences, le DD à l'interface des deux sous-unités semblent faire partie de la même famille que les DD VirA/DD VirFG (K.J. Weissman, non publié). La sous-unité 5924 est formée par deux domaines KS (en N et C-terminale), d'un KR et d'un ACP.

Caractérisation du couple de DD entre VirA et VirFG, découverte d'un nouveau type d'interface de docking

La continuité de la synthèse des polykétides est assurée par les domaines de docking qui régissent les interactions les différentes sous-unités de la PKS. Jusqu'alors, deux classes de domaines de docking avaient été caractérisée chez les PKS de type *cis*-AT (Broadhurst, et al., 2003 ; Buchholz, et al., 2009 ; Whicher, et al., 2013) et une classe chez les systèmes mixtes NRPS-PKS (Richter, et al., 2008).

Les travaux précédents sur le module 5 ont permis d'identifier un couple potentiel de DD. Etant donné que les travaux sur les interfaces de docking ont été exclusivement

menés sur les PKS de type *cis*, nous avons voulu caractériser le premier couple de DD chez une PKS de type *trans*-AT.

Lors de la purification du DD VirA, nous avons constaté que son volume d'élution correspondait à celui d'un dimère. Cet état oligomérique a été confirmé par SAXS, grâce aux calculs d'enveloppe, mais aussi grâce au programme SAXS MoW (Fischer, *et al.*, 2010). En effet, ce logiciel permet de déterminer le poids moléculaire de l'espèce présente en solution grâce à l'intégration de l'aire sous la courbe de Kratky. Nous avons également mesuré la masse apparente du DD VirA par diffusion statique de la lumière (SLS), ce qui a permis de confirmer que le DD VirA est bien dimérique en solution. Cependant lors des expériences ^{15}N - ^1H HSQC sur le DD VirA marqué ^{15}N seul en solution, seul 15 pics sur 37 attendus ont pu être identifiés. En revanche la totalité des pics apparaissent lorsque le DD VirFG est ajouté lors de l'acquisition. Deux possibilités pouvaient expliquer ce phénomène : soit il existe un équilibre entre la forme monomérique et dimérique du DD VirA et les valeurs de k_{on} et k_{off} sont trop élevées, ce qui explique la disparition des pics, ou bien il existe plusieurs formes dimériques du DD VirA. Dans ce dernier cas, les résidus subissent des environnements multiples durant l'acquisition, ce qui expliquerait la faible proportion de pics visible sur l'HSQC. Plusieurs conformations du DD VirA sous forme dimérique ont pu être mises en évidence par SAXS, ce qui appuie alors notre seconde hypothèse

Un spectre ^{15}N - ^1H HSQC a également été enregistré sur le DD VirFG. Bien que tous les pics attendus soient visibles, ceux-ci sont ramassés au centre du spectre, ce qui est caractéristique d'une protéine non-structurée. Ce résultat a d'ailleurs été conforté par une mesure de dichroïsme circulaire, qui dénote une faible structuration du DD VirFG. Ceci est aussi confirmé par l'allure de la courbe de Kratky, qui reste parallèle à l'axe des abscisses sans jamais retomber à une valeur nulle, ce qui est typique des protéines non-structurées.

Nous avons par la suite obtenue des données suggérant que le DD VirFG constitue une protéine intrinsèquement désordonnée. Des expériences de dichroïsme circulaire en double cuve ont permis de mettre en évidence une structuration du DD VirFG en présence de son partenaire voire également une structuration du DD VirA. De même, l'acquisition d'un spectre ^{15}N - ^1H HSQC du DD VirFG marqué ^{15}N a révélé un changement de déplacement chimiques des pics en présence du DD VirA ce qui

révèle une interaction. Ces pics sont alors davantage dispersés dans le spectre, ce qui démontre une structuration du DD VirFG.

Nous avons ensuite résolu par RMN la structure d'une protéine de fusion entre les deux DD. La protéine de fusion a été produite après avoir essayé de résoudre la structure d'un DD marqué ^{13}C - ^{15}N en présence de son partenaire non marqué, une approche couronnée de succès pour le DD de VirA mais qui s'est révélée trop complexe pour le DD de VirFG. Cette structure révèle une structuration en deux hélices α pour chaque DD. Les DD en interaction forment donc un fagot de 4 hélices α dans lequel les deux DD interagissent de manière perpendiculaire (un angle de 127° est formé entre les deux sets d'hélices). L'interface de docking étant composée d'une seule copie de chaque DD, et les PKS étant dimériques, il y aurait donc deux interfaces de docking distinctes.

L'interaction est principalement maintenue par des interactions de type hydrophobe et est dirigée par trois couples de résidus polaires assurant la spécificité. Une constante de dissociation (K_d) de $5,8 \mu\text{M}$ a été déterminée pour l'interaction entre le DD VirA et le DD VirFG sauvages. Cette valeur se situe dans la gamme des K_d obtenues pour les deux autres classes de DD (2 – $50 \mu\text{M}$) (Broadhurst, *et al.*, 2003 ; Buchholz, *et al.*, 2009 ; Whicher, *et al.*, 2013). Bien que le mode d'interaction varie entre les différentes classes de DD, la surface d'interaction (1100 – 1800 \AA^2) et l'affinité (2 – $50 \mu\text{M}$) reste sensiblement la même. Cette affinité est cohérente avec la concentration intracellulaire en PKS qui serait de la centaine de nanomolaires (Straight, *et al.*, 2007). Ceci permet aux sous-unités PKS d'interagir entre elles de manières spécifiques mais de façon labile. Cela suggère une habileté de la PKS à pouvoir remplacer une sous-unité défailante par une autre sans avoir à synthétiser *de novo* la machinerie entière, ce qui serait très coûteux en énergie pour la cellule. La variété entre les différents types de DD permettrait également à la PKS d'assurer des interactions très spécifiques entre les sous-unités

La présence des domaines flanquants n'affectent pas cette interaction puisque un K_d de $4 \mu\text{M}$ a été déterminé entre l'ACP_{5b}-DD VirA / DD VirFG et de $5,5 \mu\text{M}$ entre le DD VirA / DD VirFG + *linker* le reliant au KS₆. Nous n'avons en revanche pas pu tester l'influence du KS₆ sur l'interaction, étant donné que DD VirFG–KS₆ n'a jamais pu être obtenu dans la fraction soluble.

Des analyses par SAXS des constructions incorporant les domaines adjacents avaient pour but de déterminer la position relative de ces domaines vis-à-vis des DD, mais aussi de confirmer l'apparente absence de contribution à l'interaction du KS₆ et de l'ACP_{5b}. L'analyse par SAXS de la construction ACP_{5b}-DD VirA-DD VirFG a montré que l'ACP_{5b} n'interagit pas avec le couple de DD. L'étude de la protéine de fusion ACP_{5b}-DD VirA -DD VirFG-KS₆ a tout d'abord indiqué que le linker séparant le DD VirFG et le KS₆ semble structuré. Nous avons également positionné les deux domaines adjacents grâce à cette construction. Les deux DD en interaction sont flanqués d'un côté par l'ACP et de l'autre par le domaine KS. Il est intéressant de constater qu'au contraire des domaines de docking identifiés à ce jour (Broadhurst, *et al.*, 2003 ; Buchholz, *et al.*, 2009 ; Whicher, *et al.*, 2013), cette interface ne se situe pas au niveau de l'axe de symétrie du dimère de KS mais sur les flancs de ce dernier.

Les calculs de forme de la protéine de fusion ACP_{5b}-DD VirA-DD VirFG-KS₆ et du module 5 nous ont permis de proposer un modèle d'interaction entre le module 5 et le module 6. Dans ce modèle, comme les DD sont situés sur les côtés du KS, une chambre réactionnelle est formée. Ceci laisse la possibilité aux protéines de la cassette de β -méthylation d'agir tour-à-tour au sein de cette chambre réactionnelle (bien que le complexe VirA-VirFG soit transitoire) ou bien de former un large complexe macromoléculaire qui pourrait offrir une meilleure protection des intermédiaires réactifs.

Cette étude a ainsi permis de caractériser une nouvelle classe de DD, qui au vu des analyses de séquences semblent répandue au sein des PKS de type *trans*-AT. Récemment, quatre couples de DD appartenant au même type que VirA/VirFG ont été mis en évidence chez la PKS synthétisant la macrolactine (Zeng, *et al.*, 2016). Dans cette PKS, plusieurs modules sont scindées sur deux chaînes peptidiques différentes qui se complémentent en *trans*. Parmi ces quatre couples de DD, trois sont positionnés à l'interface entre DH et KR et un entre KS et KR. La structure du domaine KR possédant en N-terminale un DD (MlnE DD) permettant l'interaction avec un DD partenaire (MlnD DD) en C-terminale du domaine KS a été résolue par cristallographie à 1,72 Å de résolution (Figure 101). Le MlnE DD, analogue au DD VirFG, est structuré malgré l'absence du DD partenaire. La structure du MlnE DD est identique à celle du DD VirFG dans la protéine de fusion. Des interactions impliquent la deuxième hélice du MlnE DD avec des résidus du *linker* reliant le DD au KR et avec des résidus du

domaine KR. Ces interactions n'expliquent cependant pas à elles seules la structuration du MlnE DD. En effet, en faisant apparaître la symétrie cristalline, nous voyons qu'une autre copie du MlnE DD entre en interaction avec celui-ci (analyse non-publiée). Ce symétrique mime la présence du DD partenaire et induit vraisemblablement le repliement du MlnE DD. En revanche même s'il est structuré, celui-ci est tout de même très dynamique. Ceci est reflété par la valeur du facteur B moyen, bien supérieur pour le MlnE DD par rapport au reste de la protéine (71 contre 37 Å²).

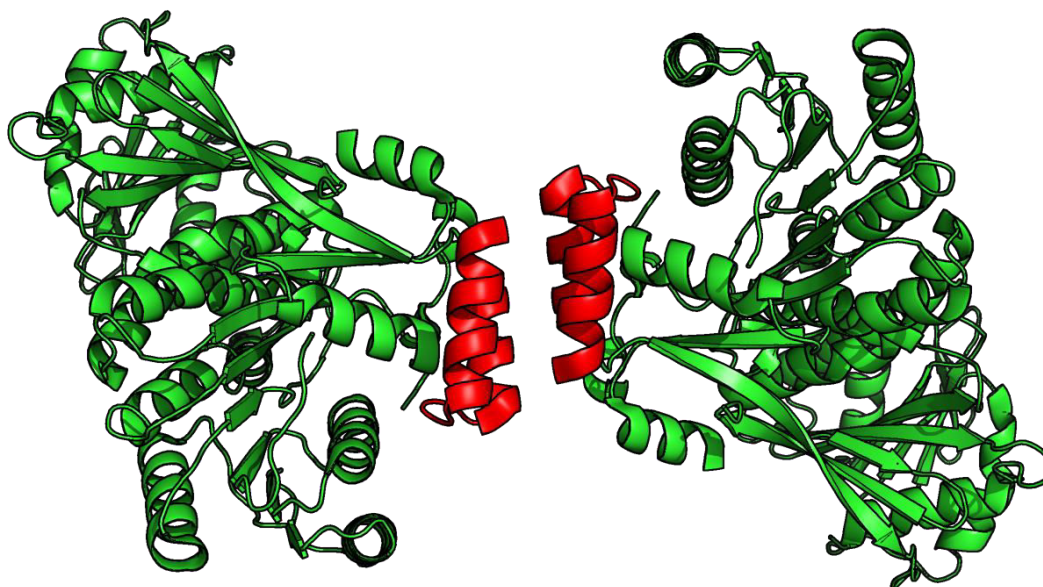


Figure 101 : Structure d'un DD-KR de la sous-unité MlnE de la PKS synthétisant la macrolactine (PDB 5D2E, Zeng, et al., 2016).

Le domaine KR est coloré en vert, et le domaine MlnE DD en rouge. Un symétrique est affiché ici. Nous voyons que dans cette structure, le DD s'homodimérise, ce qui explique probablement la structuration de celui-ci. Cette interaction entre DD s'effectue en revanche avec les deux sets d'hélices qui s'organisent de façon parallèle contrairement au couple entre les DD VirA et DD VirFG pour lesquels les deux sets d'hélices forment un angle de 127°.

L'interface de docking entre VirA et VirFG est totalement différente de celles décrites jusqu'à présent, que ce soit au niveau de l'état oligomérique ou bien de la localisation par rapport au domaine KS en aval. Cela constitue une avancée majeure pour l'ingénierie de ces systèmes enzymatiques. Nous avons identifié des résidus polaires importants pour la spécificité et la mise en place de l'interaction (k_{on}). Ces résidus peuvent alors être la cible de mutagenèse dirigée dans le but de modifier la spécificité d'interaction. En revanche, du fait du caractère intrinsèquement désordonné

de certains, il est impossible de prévoir si une mutation va affecter la structure ou non. Un échange complet de ces domaines est donc à privilégier afin de maintenir leur intégrité structurale. La faisabilité de cet échange a d'ailleurs été démontrée lors de l'étude sur les DD de la PKS synthétisant la macrolactine (Zeng, *et al.*, 2016). Cependant, il est également très important de prendre garde à échanger ces DD par un couple de DD n'interagissant pas avec des DD déjà présent au sein de la PKS. Afin d'éviter ce problème, il est nécessaire de caractériser plusieurs interfaces de docking supplémentaires, un travail actuellement en cours au laboratoire, qui constitue une partie du sujet de thèse de Fanny Risser.

Etude d'une enzyme de cyclisation post-PKS

Les enzymes post-PKS représentent une part très importante de la synthèse du polykétide. En effet, comme nous avons pu le voir dans l'introduction, ces enzymes sont responsables de modifications du polykétide étant essentielles pour sa fonction. Le choix de la LkcE comme cible d'étude présente un intérêt double puisqu'il s'agit à la fois d'une enzyme post-PKS mais aussi d'une enzyme d'oxydation et de cyclisation. Ainsi la compréhension du mécanisme catalytique de l'enzyme permettrait de disposer d'un outil supplémentaire pour l'ingénierie des PKS (en cas de spécificité de substrat suffisamment large ou altérable) mais aussi de détenir une enzyme capable de cycliser des molécules, réaction qui est très difficile à réaliser en chimie traditionnelle (Ojima, 2002).

Nous avons donc procédé à une étude structurale de l'enzyme, en essayant de résoudre sa structure en présence d'analogues de substrat, ce qui nous a permis d'identifier des résidus pouvant être impliqués dans la catalyse. En effet, outre le FAD qui va permettre l'oxydation de l'amine de l'amide en imine, et ainsi la formation d'un centre électrophile, il semble nécessaire qu'une catalyse base puisse avoir lieu pour activer le C2 du polykétide. Une attaque nucléophile intramoléculaire doit ensuite avoir lieu pour aboutir à un produit cyclique.

La structure de la LkcE adopte un repliement classique d'amine oxydase. En effet, la protéine est composée d'un domaine de fixation du FAD et d'un domaine catalytique. L'ensemble se dimérise par l'intermédiaire d'une large interface entre les deux

monomères. Le repliement global est semblable à celui d'autres amines oxydases, particulièrement au niveau du domaine de fixation du FAD. Ce dernier présente également une conservation de la séquence primaire plus importante que le reste de la protéine avec les autres amines oxydases (Figure 72).

Grâce aux premières structures (SeLkcE et LkcE_{wt} avec analogues de substrat), nous avons pu identifier la Y182 ainsi que la R326 comme des résidus pouvant être impliqués dans la catalyse. En effet, peu de résidus présent dans le site actif de l'enzyme sont capables de jouer le rôle de catalyseur base et sont conservés, notamment avec la LkcE de *P. fluorescens*. Seule la Y182 semble pouvoir jouer ce rôle, bien que le pK_a de ce résidu doive être fortement diminué par son environnement (environ 3 unité de pH). La R326 est quant à elle, située en fond du site actif et nous avons alors supposé qu'elle puisse servir dans la stabilisation de l'énolate lors de la catalyse. Par ailleurs, dans la structure de la LkcE_{wt} avec les analogues de substrat, on peut remarquer que le DATD n'est pas positionné de façon à subir une oxydation par le FAD. En effet la fonction amide du DATD est située à une distance de 6,7 Å de l'azote réactif du FAD. Cependant, on constate que cette fonction amide établit une liaison hydrogène avec la T397 qui pourrait dans ce cas stabiliser l'imine formé lors de la catalyse avec le substrat naturel de l'enzyme. Cela nécessiterait néanmoins un changement conformationnel important de la part de ce dernier. .

Au sein de l'autre monomère, la fixation de l'EMAA au cœur du site actif ne fournit aucun renseignement sur la catalyse, puisqu'aucun résidu potentiellement catalytique n'est situé proche ou orienté correctement pour permettre une catalyse base sur le carbone de l'EMAA mimant le C2 du polykétide.

Une fois le substrat naturel de la LkcE purifié par les chimistes du laboratoire du Pr. A. Kirschning à hauteur de 20 mg pour 20 litres de cultures extraits à l'acétate d'éthyle, nous avons tenté d'obtenir la structure de la LkcE avec le LC-KA05. De manière inattendue, par trempage des cristaux dans le substrat, nous sommes parvenues à obtenir le substrat dans le site actif de l'enzyme seulement pour le mutant R326Q. Le polykétide est fixé sous sa forme linéaire dans le site actif. Nous avons vu que dans cette conformation, l'amine se situe à proximité de l'azote réactif du groupement isoalloxazine du FAD. Ce positionnement est idéal pour la catalyse de l'oxydation de l'amine en imine.

En revanche, le fait que le substrat soit linéaire dans le site actif de la LkcE n'est pas favorable à la réaction de cyclisation. En effet le C2 est séparé de la fonction amine de 13,2 Å. De plus, dans cet état, le C2 se situe à l'entrée du canal du site actif et aucun résidu susceptible de réaliser une catalyse base ne se situe à proximité de ce carbone. Cet état reflète donc une partie de la catalyse, mais un réarrangement conformationnel permettant de contraindre le substrat semble nécessaire au bon déroulement de la réaction de cyclisation. Dans ce canal, peu de résidus « candidats » à une catalyse basique sont présents. Seul une histidine (H65), un glutamate (E64) situés en fond de site actif et une tyrosine (Y182) sont des candidats potentiels. Seulement, la H65 n'est pas conservée chez les LkcE des systèmes homologues permettant la synthèse de la lankacidine. Seul la Y182 ou la E64 resteraient de bons candidats pour jouer le rôle de catalyseur base. Néanmoins, la Y182 est également conservée parmi les MAO humaines qui ne catalysent pas de réactions de déprotonation. Ceci pourrait indiquer que la Y182 ne serait pas impliquée dans une catalyse basique.

Un autre point important est que seul le mutant R326Q a permis l'obtention du substrat dans le site actif. La Q326 établit une liaison hydrogène avec le LC-KA05 contribuant à la fixation du substrat. Cependant, un point remarquable est que la R326 comme positionnée dans la structure en complexe avec les analogues de substrat, entrerait en clash stérique avec le LC-KA05 (Figure 102). Cela démontre que le site actif s'accommoderait lors de la fixation du substrat.

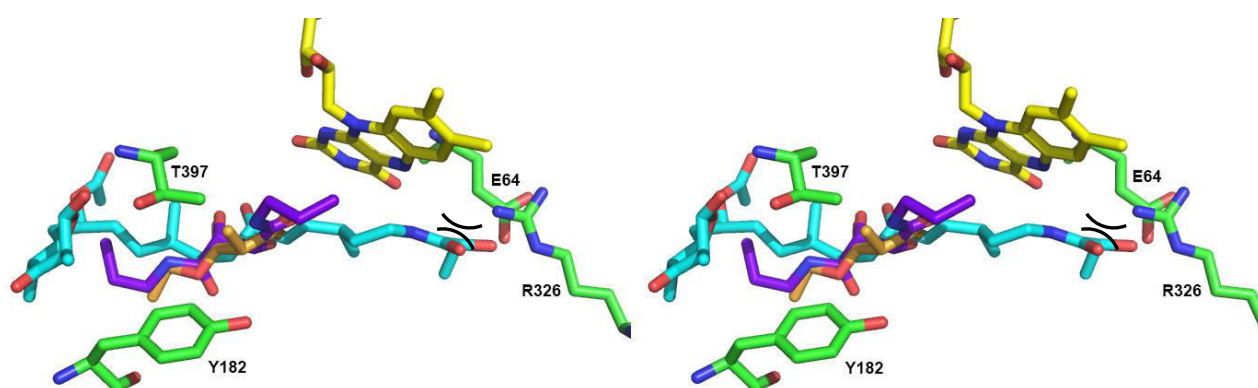


Figure 102 : Vue stéréo du site actif de la LkcE_{wt}.

Les structures de la LkcE_{wt} (en vert) et du mutant R326Q ont été superposées, permettant d'identifier le site de fixation du DATD (en mauve) et de l'EMAA (en orange) par rapport au substrat naturel de l'enzyme (en cyan). Dans cette conformation, la R326 entrerait en clash stérique (indiqué par des traits noirs) avec le LC-KA05 (1,5 Å sépare le NH₂ de la R326 et l'hydroxyle du substrat).

La structure du mutant R326Q en présence du substrat reflète certainement la première étape de la catalyse. En effet, la fonction amine est située de façon adéquate par rapport au FAD pour que le transfert d'électron puisse s'effectuer. Dans cette organisation, la réaction ne peut aller plus loin, d'une part parce que le C2 et le C18 sont situés à 13 Å l'un de l'autre, et d'autre part aucun résidu pouvant jouer le rôle de base ne se situe à proximité du C2. Un réarrangement conformationnel, probablement provoqué par la formation de l'imine, est alors nécessaire pour assurer la catalyse base au niveau du C2, étape nécessaire à la cyclisation. Une stabilisation de l'imine précédemment formé est requise durant cette étape. Pour cela, deux résidus sont potentiellement impliqués : l'E64, situés dans le fond du site actif, et la T397 située vers le milieu du canal d'entrée du substrat. Ces deux résidus sont également conservés chez les LkcE homologues.

Une analyse RMN effectuée par les chimistes allemands a par ailleurs montré que le substrat serait à 100 % sous forme énol. Cela prouve que le proton relié au C2 présente un caractère acide. En effet, un système conjugué impliquant la fonction ester en C1 rend le C2 particulièrement acide en diminuant son pK_a . Cependant un énol n'est pas aussi réactif qu'un énolate. Ainsi, afin d'accroître la réactivité du C2 et catalyser la cyclisation, une déprotonation de l'hydroxyle doit avoir lieu. Celle-ci peut impliquer l'E64 ou la Y182 qui sont les deux seuls résidus présent dans le site actif pouvant participer à une catalyse basique. La réaction de cyclisation serait alors concomitante à la déprotonation. La cyclisation aboutirait alors à l'inversion de la stéréochimie du groupement méthyle (C19).

Comme décrit précédemment, pour que la réaction de cyclisation puisse avoir lieu, il est nécessaire de contraindre le substrat au sein du site actif afin de rapprocher le C de la fonction amine. La formation de l'imine, et par conséquent l'apparition d'une charge positive, pourrait générer une répulsion entre l'imine et la R326 située à proximité. Cette répulsion pourrait alors permettre d'agencer le substrat dans une configuration correcte pour la réaction de cyclisation ou alors induire un réarrangement conformationnel de l'enzyme permettant de positionner le substrat dans une orientation correcte pour réagir. Un schéma descriptif de la réaction est proposé en Figure 103. La régénération du FAD se fait ensuite par une molécule d'O₂ qui est alors réduite en H₂O₂.

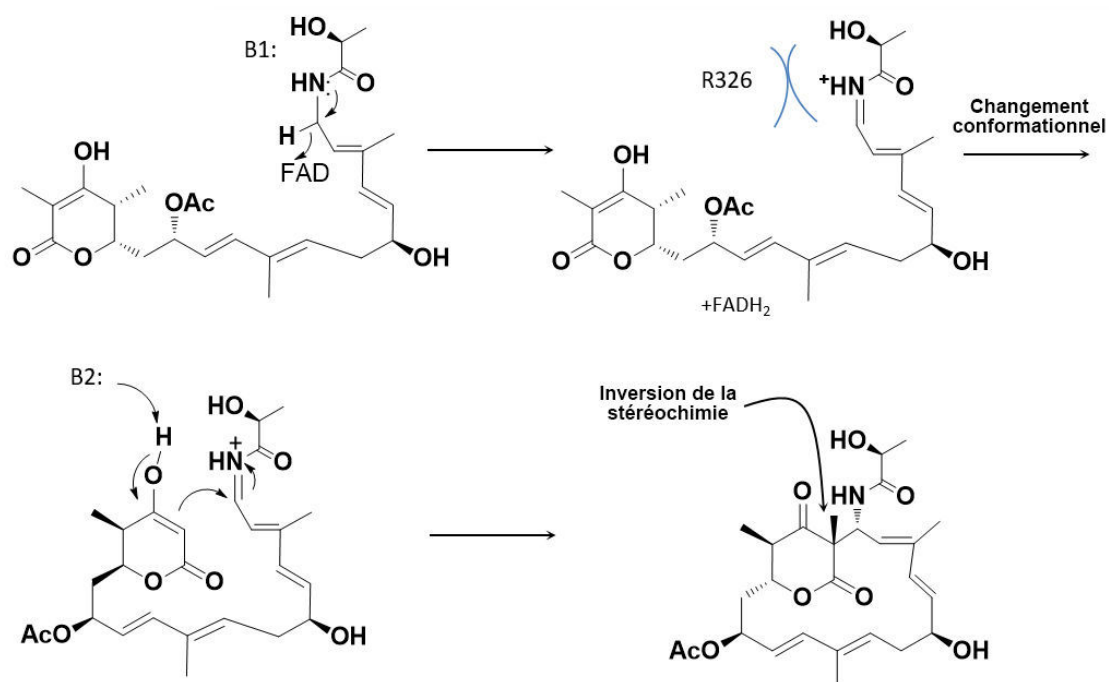


Figure 103 : Schéma réactionnel proposé pour la LkcE (réalisée par KJ Weissman).

Le FAD serait responsable de l'oxydation de l'amine en imine, grâce à un transfert d'hydrure. L'imine ainsi formée serait stabilisée par un résidu B1 (le E64 pourrait stabiliser l'intermédiaire). La formation de cette imine induirait un changement conformationnel qui permettrait de positionner le substrat dans une conformation adéquate pour la catalyse. Dès lors, un résidu B2 (le E64 est le candidat le plus probable, la Y182 pourrait également jouer ce rôle si son pK_a est diminué) agirait par catalyse base pour activer le substrat sous forme énol. La cyclisation serait alors concomitante à l'activation, ce qui conduit à une inversion de stéréochimie.

Le suivi de la réaction avec les analogues de substrat s'est révélé infructueux. En effet, la réaction ne semble pas avoir lieu en présence de DATD et d'EMAA. Ceci n'est guère étonnant au vu des structures, puisque les deux analogues de substrat se fixent au même endroit dans le site actif. Le site de fixation des deux analogues est par ailleurs assez étonnant puisqu'ils se positionnent au niveau de la chaîne aliphatique du LC-KA05 (Figure 102).

Le suivi par LC-MS de la réaction avec le substrat naturel de l'enzyme s'est révélé être plus compliqué que prévu. D'une part, la gamme étalon permettant de quantifier le substrat n'est pas linéaire dans les concentrations utilisées pour le suivi de l'activité. De plus, jusqu'à maintenant, il a été impossible de mettre en évidence par cette méthode le produit formé. Une méthode de suivi de la réaction par mesure de la concentration en O₂ pourrait donc fournir des résultats plus rigoureux. Cette méthode a déjà été utilisée pour suivre l'activité d'une amine oxydase (Ames, *et al.*, 2011). De plus, la diffusion de l'O₂ dans l'eau est suffisamment lente pour ne pas interférer avec

le suivi de la consommation d'O₂ par l'enzyme. En effet, cette vitesse est de l'ordre de 10⁻⁵ cm².s⁻¹ (Han, et Bartels, 1996).

Ainsi, afin d'envisager une publication de ce travail, il est critique de déterminer les paramètres cinétiques de l'enzyme, ce qui peut être fait en utilisant une électrode O₂, mais aussi de mettre en évidence la formation du produit. Pour cela, nous disposons de deux possibilités : 1) réaliser une extraction de la souche sauvage de *S. rochei*, ce qui nous permettra de disposer d'un témoin pour analyser la présence du produit par LC-MS (l'extraction de la souche sauvage contient en effet le produit de la réaction catalysée par la LkcE). Cela permettrait en effet de déterminer les conditions optimales pour mettre en évidence le produit de la réaction. 2) Après avoir stoppé la réaction et extrait le produit à l'acétate d'éthyle, celui-ci pourrait être analysé en RMN. Le produit se dégradant dans une solution aqueuse, il serait préférable d'enregistrer les données RMN en reprenant le produit dans du chloroforme deutéré.

REFERENCES
BIBLIOGRAPHIQUES

1. Agostini-costa, T. S., Vieira, R. F., Bizzo, H. R., Silveira, D. & Gimenes, M. a. Secondary Metabolites. *Chromatogr. Its Appl.* **1**, 131–164 (2012).
2. Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M. & Thorson, J. S. The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173–1176 (2002).
3. Akey, D. L., Kittendorf, J. D., Giraldes, J. W., Fecik, R. A., Sherman, D. H. & Smith, J. L. Structural basis for macrolactonization by the pikromycin thioesterase. *Nat. Chem. Biol.* **2**, 537–542 (2006).
4. Akey, D. L., Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H. & Smith, J. L. Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure* **18**, 94–105 (2010).
5. Alekseyev, V. Y., Liu, C. W., Cane, D. E. & Puglisi, J. D. Solution structure and proposed domain – domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16**, 2093–2107 (2007).
6. Ames, B. D., Nguyen, C., Bruegger, J., Smith, P., Xu, W., Ma, S., Wong, E., Wong, S., Xie, X., Li, J. W.-H., Vederas, J. C., Tang, Y. & Tsai, S.-C. Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11144–9 (2012).
7. Ames, B. D., Haynes, S. W., Gao, X., Evans, B. S., Kelleher, N. L., Tang, Y. & Walsh, C. T. Oxidation of Fumiquinazoline A to the Heptacyclic Hemiaminal Fumiquinazoline C by the Flavoenzyme Af12070 from *Aspergillus fumigatus*. *Biol. Chem.* **50**, 8756–8769 (2011).
8. Annaval, 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).

9. Arakawa, K., Mochizuki, S., Yamada, K., Noma, T. & Kinashi, H. γ -Butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei*. *Microbiology* **153**, 1817–1827 (2007).
10. 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**, 249–56 (2005).
11. Arthur, C. J., Szafranska, A., Evans, S. E., Findlow, S. C., Burston, S. G., Owen, P., Clark-Lewis, I., Simpson, T. J., Crosby, J. & Crump, M. P. Self-malonylation is an intrinsic property of a chemically synthesized type II polyketide synthase acyl carrier protein. *Biochemistry* **44**, 15414–15421 (2005).
12. Audic, S., Lopez, F., Claverie, J. M., Poirot, O. & Abergel, C. SAmBA: An interactive software for optimizing the design of biological macromolecules crystallization experiments. *Proteins Struct. Funct. Genet.* **29**, 252–257 (1997).
13. Auerbach, T., Mermershtain, I., Davidovich, C., Bashan, A., Belousoff, M., Wekselman, I., Zimmerman, E., Xiong, L., Klepacki, D., Arakawa, K., Kinashi, H., Mankin, A. S. & Yonath, A. The structure of ribosome-lankacidin complex reveals ribosomal sites for synergistic antibiotics. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1983–8 (2010).
14. Bachmann, B. O., Van Lanen, S. G. & Baltz, R. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **41**, 175–184 (2014).
15. Bachmann, B. O., Van Lanen, S. G. & Baltz, R. H. Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **41**, 175–184 (2014).
16. Bachmann, B. O. & Ravel, J. Chapter 8 Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. *Methods Enzymol.* **458**, 181–217 (2009).

17. Belousoff, M. J., Shapira, T., Bashan, A., Zimmerman, E., Rozenberg, H., Arakawa, K., Kinashi, H. & Yonath, A. Crystal structure of the synergistic antibiotic pair, lankamycin and lankacidin, in complex with the large ribosomal subunit. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2717–22 (2011).
18. Bevitt, D. J., Staunton, J. & Leadlay, P. F. Mutagenesis of the dehydratase active site in the erythromycin-producing polyketide synthase. *Biochem. Soc. Trans.* **21**, 30S (1993).
19. Binda, C., Angelini, R., Federico, R., Ascenzi, P. & Mattevi, A. Structural bases for inhibitor binding and catalysis in polyamine oxidase. *Biochemistry* **40**, 2766–2776 (2001).
20. Bisang, C., Long, P. F., Cortés, J., Westcott, J., Crosby, J., Matharu, a L., Cox, R. J., Simpson, T. J., Staunton, J. & Leadlay, P. F. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**, 502–505 (1999).
21. Blundell, T. L. & Johnson, L. N. *Protein Crystallography*. (Academic Press, New York, 1976).
22. Bravo-Rodriguez, K., Ismail-Ali, A. F., Klopries, S., Kushnir, S., Ismail, S., Fansa, E. K., Wittinghofer, A., Schulz, F. & Sanchez-Garcia, E. Predicted incorporation of non-native substrates by a polyketide synthase yields bioactive natural product derivatives. *ChemBiochem* **15**, 1991–1997 (2014).
23. Bretschneider, T., Heim, J. B., Heine, D., Winkler, R., Busch, B., Kusebauch, B., Stehle, T., Zocher, G. & Hertweck, C. Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124–8 (2013).
24. Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat. Struct. Mol. Biol.* **16**, 190–7 (2009).
25. 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).

26. Buchholz, T. J., Geders, T. W., Bartley III, F. E., Reynolds, K. A., Smith, J. L. & Sherman, D. H. Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
27. Burmester, J., Haese, A. & Zocher, R. Highly conserved N-methyltransferases as an integral part of peptide synthetases. *Biochem. Mol. Biol. Int.* **37**, 201–7 (1995).
28. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* **4**, 654–657 (2003).
29. Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* **25**, 845–53 (2008).
30. Calderone, C. T., Iwig, D. F., Dorrestein, P. C., Kelleher, N. L. & Walsh, C. T. Incorporation of Nonmethyl Branches by Isoprenoid-like Logic: Multiple β -Alkylation Events in the Biosynthesis of Myxovirescin A1. *Chem. Biol.* **14**, 835–846 (2007).
31. Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T. & Dorrestein, P. C. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8977–82 (2006).
32. Carter, C. W. & Carter, C. W. Protein Crystallization Using Incomplete Factorial-Experiments. *J. Biol. Chem.* **254**, 2219–2223 (1979).
33. Chakravarty, B., Gu, Z., Chirala, S. S., Wakil, S. J. & Quioco, F. A. Human fatty acid synthase: structure and substrate selectivity of the thioesterase domain. *Proc Natl Acad Sci U S A* **101**, 15567–15572 (2004).
34. Challis, G. L. & Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Struct. Biol.* **14**, 748–756 (2004).

35. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
36. Cheng, Y., Tang, G. & Shen, B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–3154 (2003).
37. Cocito, B. C. Metabolism of Macromolecules in Bacteria Treated with Virginiamycin. *J. Gen. Microbiol.* **57**, 179–194 (1969).
38. Cocito, C. Antibiotics of the virginiamycin family, inhibitors which contain synergistic components. *Microbiol. Rev.* **43**, 145–192 (1979).
39. De Colibus, L., Li, M., Binda, C., Lustig, A., Edmondson, D. E. & Mattevi, A. Three-dimensional structure of human monoamine oxidase A (MAO A): relation to the structures of rat MAO A and human MAO B. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12684–9 (2005).
40. Collie, J. N. & Myers, W. S. The Formation of Orcinol and other Condensation Products from Dehydracetic acid. *J. Chem. Soc.* **63**, 122–128 (1893).
41. Combet, C., Blanchet, C., Geourjon, C. & Del?age, G. NPS@: Network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150 (2000).
42. Cortés, 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).
43. Cowtan, K. D. & Zhang, K. Y. J. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72**, 245–270 (1999).
44. Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5**, 2010–26 (2007).

45. 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 (2009).
46. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. & Richardson, D. C. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, 375–383 (2007).
47. 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).
48. Demain, A. L. & Fang, A. The natural functions of secondary metabolites. *Adv. Biochem. Eng. Biotechnol.* **69**, 222 (2000).
49. Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biotechnol.* **41**, 185–201 (2014).
50. Desai, R. P., Rodriguez, E., Galazzo, J. L. & Licari, P. Improved bioconversion of 15-fluoro-6-deoxyerythronolide B to 15-fluoro-erythromycin A by overexpression of the eryK gene in *Saccharopolyspora erythraea*. *Biotechnol. Prog.* **20**, 1660–1665 (2004).
51. Dickschat, J. S., Vergnolle, O., Hong, H., Garner, S., Bidgood, S. R., Dooley, H. C., Deng, Z., Leadlay, P. F. & Sun, Y. An additional dehydratase-like activity is required for Lankacidin antibiotic biosynthesis. *ChemBioChem* **12**, 2408–2412 (2011).
52. Dillon, S. C. & Bateman, A. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC Bioinf.* **5**, 109 (2004).
53. 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**, 7119–23 (1993).

54. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **252**, 675–679 (1991).
55. Du, L., Edwards, D. J., Toney, M. D. & Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide- polyketide antitumor drug bleomycin. **8**, 725–738 (2001).
56. Ducruix, A. & Giegé, R. *Crystallization of Nucleic Acids and Proteins. A Practical Approach. Molecular Biology* (1999).
57. Duke, E. M. H. & Johnson, L. N. Macromolecular crystallography at synchrotron radiation sources: current status and future developments. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **466**, 3421–3452 (2010).
58. Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P. & Fieschi, F. NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53 (2010).
59. 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. & Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
60. Dym, O. & Eisenberg, D. Sequence-structure analysis of FAD-containing proteins. 1712–1728 (2001). doi:10.1101/ps.12801.structure
61. Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. Architectures of Whole-Module and Bimodular Proteins from the 6-Deoxyerythronolide B Synthase. *J. Mol. Biol.* **426**, 2229–2245 (2014).
62. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
63. Evans, B. S., Chen, Y., Metcalf, W. W., Zhao, H. & Kelleher, N. L. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. *Chem. Biol.* **18**, 601–607 (2011).

64. Evans, G. A., Lewis, K. & Rothenberg, B. E. High efficiency vectors for eosmid microcloning and genomic analysis (Recombinant DNA; gene libraries; bacteriophage I cos site; promoters of phages T3 and T7; mammalian chromosomes). **79**, 9–20 (1989).
65. Fagan, R. L. & Palfey, B. A. in *Comprehensive Natural Products II* (ed. Begley, T. P.) **7**, 37–114 (2010).
66. Fernandez-Moreno, M. A., Martinez, E., Boto, L., Hopwood, D. A. & Malpartida, F. Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3 (2) including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem.* **267**, 19278–19290 (1992).
67. Fichtlscherer, F., Wellein, C., Mittag, M. & Schweizer, E. A novel function of yeast fatty acid synthase Subunit a is capable of self-pantetheinylation. *Eur. J. Biochem.* **2671**, 2666–2671 (2000).
68. Fischbach, M. a & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–96 (2006).
69. Fischer, H., Neto, M. D. O., Napolitano, H. B. & Polikarpov, I. 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).
70. Fitzpatrick, P. F. Oxidation of amines by flavoproteins. *Arch. Biochem. Biophys.* **493**, 13–25 (2010).
71. Flavin, R., Peluso, S., Nguyen, P. & Loda, M. Fatty acid synthase as a potential therapeutic target in cancer. *Futur. Oncol.* **6**, 551–562 (2010).
72. Franke, D., Jeffries, C. M. & Svergun, D. I. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12**, 419–422 (2015).
73. Gaisser, S., Lill, R., Staunton, J., Méndez, C., Salas, J. & Leadlay, P. F. Parallel pathways for oxidation of 14-membered polyketide macrolactones in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **44**, 771–781 (2002).

74. Garg, A., Khosla, C. & Cane, D. E. Coupled methyl group epimerization and reduction by polyketide synthase ketoreductase domains. Ketoreductase-catalyzed equilibrium isotope exchange. *J. Am. Chem. Soc.* **135**, 16324–16327 (2013).
75. 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**, 10190–10193 (2014).
76. Gay, D. C., Wagner, D. T., Meinke, J. L., Zogzas, C. E., Gay, G. R. & Keatinge-Clay, A. T. The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. *J. Struct. Biol.* **193**, 196–205 (2016).
77. Gay, D., You, Y. O., Keatinge-Clay, A. & Cane, D. E. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry* **52**, 8916–8928 (2013).
78. Gehring, A. M., Mori, I., Perry, R. D. & Walsh, C. T. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**, 11637–11650 (1998).
79. Girardot, R., Viguier, G., Pérez, J. & Ounsy, M. FOXTROT: A Java-based application to reduce and analyse SAXS and WAXS piles of 2D data at synchrotron SOLEIL. *Abstr. du J-PARC, Tokai, Japan* (2011).
80. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
81. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).

82. Gu, L., Eisman, E. B., Dutta, S., Franzmann, T. M., Walter, S., Gerwick, W. H., Skiniotis, G. & Sherman, D. H. Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect. *Angew. Chemie - Int. Ed.* **50**, 2795–2798 (2011).
83. Han, P. & Bartels, D. M. Temperature Dependence of Oxygen Diffusion in H₂O and D₂O. *J. Phys. Chem.* **100**, 5597–5602 (1996).
84. Hansen, J. L., Moore, P. B. & Steitz, T. a. Structures of Five Antibiotics Bound at the Peptidyl Transferase Center of the Large Ribosomal Subunit. *J. Mol. Biol.* **330**, 1061–1075 (2003).
85. Hayashi, T., Suenaga, I., Narukawa, N. & Yamazaki, T. In vitro and in vivo activities of sedecamycin against *Treponema hyodysenteriae*. *Antimicrob. Agents Chemother.* **32**, 458–461 (1988).
86. Haydock, S. F., Aparicio, J. F., König, A., Molnár, I., Schwecke, T., Marsden, A. F. A., Galloway, I. S. M., Staunton, J. & Leadlay, P. F. Divergent structural motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374**, 246–248 (1995).
87. He, J., Müller, M. & Hertweck, C. Formation of the aureothin tetrahydrofuran ring by a bifunctional cytochrome P450 monooxygenase. *J. Am. Chem. Soc.* **126**, 16742–16743 (2004).
88. 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).
89. Hendrickson, W. a, Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).
90. Herbst, D. A., Jakob, R. P., Zähringer, F. & Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **531**, 533–537 (2016).

91. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–716 (2009).
92. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162–190 (2007).
93. 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).
94. Holm, L. & Rosenström, P. Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, 545–549 (2010).
95. Hong, H., Stinear, T., Skelton, P., Spencer, J. B. & Leadlay, P. F. Structure elucidation of a novel family of mycolactone toxins from the frog pathogen *Mycobacterium* sp. MU128FXT by mass spectrometry. *Chem. Commun. (Camb)*. 4306–4308 (2005). doi:10.1039/b506835e
96. Jancarik, J. & Kim, S. H. Sparse matrix sampling. A screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411 (1991).
97. Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **22**, 2027–2039 (2005).
98. Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolásek, B. & Ban. Structure of Fungal Fatty Acid Synthase and Implications for Iterative Substrate Shuttling. *Science* **316**, (2007).
99. Johnston, N. J., Mukhtar, T. a & Wright, G. D. Streptogramin antibiotics: mode of action and resistance. *Curr. Drug Targets* **3**, 335–344 (2002).
100. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
101. Kachalova, G., Decker, K., Holt, A. & Bartunik, H. D. Crystallographic snapshots of the complete reaction cycle of nicotine degradation by an amine oxidase of the monoamine oxidase (MAO) family. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4800–4805 (2011).

102. Kanne, R. & Zahner, H. Comparative studies on intracellular potassium- and sodium concentrations of wild-type and a macrotetrolide negative mutant of *Streptomyces griseus*. *Z. Naturforsch. C.* **31**, 115–117 (1976).
103. Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **384**, 941–953 (2008).
104. Keatinge-Clay, A. T. The Structural Relationship between Iterative and Modular PKSs. *Cell Chem. Biol.* **23**, 540–542 (2016).
105. Keatinge-clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
106. Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **14**, 898–908 (2007).
107. Keatinge-Clay, A. T., Maltby, D. a, Medzihradzky, K. F., Khosla, C. & Stroud, R. M. An antibiotic factory caught in action. *Nat. Struct. Mol. Biol.* **11**, 888–93 (2004).
108. 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).
109. Kinashi, H., Mori, E., Hatani, A. & Nimi, O. Isolation and characterization of linear plasmids from lankacidin-producing *Streptomyces* species. *J. Antibiot. (Tokyo)*. **47**, 1447–1455 (1994).
110. Kittendorf, J. D. & Sherman, D. H. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **17**, 597–605 (2006).
111. Koryakina, I., McArthur, J., Randall, S., Draelos, M. M., Musiol, E. M., Muddiman, D. C., Weber, T. & Williams, G. J. Poly specific trans -Acyltransferase machinery revealed via engineered acyl-coa synthetases. *ACS Chem. Biol.* **8**, 200–208 (2013).
112. Koski, M. K., Haapalainen, A. M., Hiltunen, J. K. & Glumoff, T. A two-domain structure of one subunit explains unique features of eukaryotic hydratase 2. *J. Biol. Chem.* **279**, 24666–24672 (2004).

113. Kratky, O. & Porod, G. Röntgenuntersuchung gelöster Fadenmoleküle. *Recl. des Trav. Chim. des Pays-Bas* **68**, 1106–1122 (1949).
114. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
115. 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).
116. 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).
117. Kwan, D. H. & Leadlay, P. F. Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS Chem. Biol.* **5**, 829–838 (2010).
118. Kwon, H. J., Smith, W. C., Xiang, L. & Shen, B. Cloning and heterologous expression of the macrotetrolide biosynthetic gene cluster revealed a novel polyketide synthase that lacks an acyl carrier protein. *J. Am. Chem. Soc.* **123**, 3385–6 (2001).
119. Labonte, J. W. & Townsend, C. A. Active site comparisons and catalytic mechanisms of the hot dog superfamily. *Chem. Rev.* **113**, 2182–2204 (2013).
120. Lambalot, R. H., Gehring, a M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. a, Reid, R., Khosla, C. & Walsh, C. T. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
121. Lee, S. K., Basnet, D. B., Choi, C. Y., Sohng, J. K., Ahn, J. S. & Yoon, Y. J. The role of erythromycin C-12 hydroxylase, EryK, as a substitute for PikC hydroxylase in pikromycin biosynthesis. *Bioorg. Chem.* **32**, 549–559 (2004).
122. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**, 253–264 (1996).

123. Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase. *Science* **316**, 288–290 (2007).
124. Liang, H., Sandberg, W. S. & Terwilliger, T. C. Genetic fusion of subunits of a dimeric protein substantially enhances its stability and rate of folding. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7010–4 (1993).
125. Liavonchanka, A., Hornung, E., Feussner, I. & Rudolph, M. G. Structure and mechanism of the *Propionibacterium acnes* polyunsaturated fatty acid isomerase. *Proc. Natl. Acad. Sci. USA* **103**, 2576–2581 (2006).
126. Lomakin, I. B., Xiong, Y. & Steitz, T. A. The Crystal Structure of Yeast Fatty Acid Synthase , a Cellular Machine with Eight Active Sites Working Together. *Cell* **129**, 319–332 (2007).
127. Mahenthiralingam, E., Song, L., Sass, A., White, J., Wilmot, C., Marchbank, A., Boaisa, O., Paine, J., Knight, D. & Challis, G. L. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic burkholderia ambifaria genomic island. *Chem. Biol.* **18**, 665–677 (2011).
128. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315–22 (2008).
129. Maier, T., Leibundgut, M., Boehringer, D. & Ban, N. Structure and function of eukaryotic fatty acid synthases. *Q Rev Biophys.* **43**, 373–422 (2010).
130. Marsden, A. F., Wilkinson, B., Cortés, J., Dunster, N. J., Staunton, J. & Leadlay, P. F. Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science* **279**, 199–202 (1998).
131. Marsden, A. F. A., Caffrey, P., Aparicio, J. F., Loughran, M. S., Staunton, J., Leadlay, P. F. & Mark, S. Stereospecific Acyl Transfers on the Polyketide Synthase Erythromycin-Producing. *Science* **263**, 378–380 (1994).
132. Mast, Y., Weber, T., Gölz, M., Ort-Winklbauer, R., Gondran, A., Wohlleben, W. & Schinko, E. Characterization of the ‘pristinamycin supercluster’ of *Streptomyces pristinaespiralis*. *Microb. Biotechnol.* **4**, 192–206 (2011).

133. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
134. 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).
135. McDaniel, R., Ebert-Khosla, S., Hopwood, D. a & Khosla, C. Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits. *Nature* **375**, 549–554 (1995).
136. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr. Sect. FStructural Biol. Commun.* **70**, 2–20 (2014).
137. Menzella, H. G., Carney, J. R. & Santi, D. V. Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol.* **14**, 143–151 (2007).
138. Menzella, H. G. & Reeves, C. D. Combinatorial biosynthesis for drug development. *Curr. Opin. Microbiol.* **10**, 238–245 (2007).
139. Metz, J. G., Metz, J. G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., Valentine, R., Lardizabal, K., Domergue, F., Yamada, A., Yazawa, K., Knauf, V. & Browse, J. Production of Polyunsaturated Fatty Acids by Polyketide Synthases in Both Prokaryotes and Eukaryotes. *Science* **290**, (2010).
140. Monciardini, P., Iorio, M., Maffioli, S., Sosio, M. & Donadio, S. Discovering new bioactive molecules from microbial sources. *Microb. Biotechnol.* **7**, 209–220 (2014).
141. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255 (1997).
142. Mylonas, E. & Svergun, D. I. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Cryst* **40**, 245–249 (2007).
143. Namwat, W., Kamioka, Y., Kinoshita, H., Yamada, Y. & Nihira, T. Characterization of virginiamycin S biosynthetic genes from *Streptomyces virginiae*. *Gene* **286**, 283–290 (2002).

144. 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**, 225–33 (2008).
145. Nicholls, R. A., Long, F. & Murshudov, G. N. Low-resolution refinement tools in REFMAC5. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 404–417 (2012).
146. Nilius, A. M. & Ma, Z. Ketolides: the future of the macrolides? *Curr. Opin. Pharmacol.* **2**, 493–500 (2002).
147. Nyman, M. & Fullmer, L. *Small angle x-ray scattering of group V polyoxometalates. Trends in Polyoxometalates* (Nova Science Publishers, Inc., 2015).
148. Nyssen, E., Di Giambattista, M. & Cocito, C. Analysis of the reversible binding of virginiamycin M to ribosome and particle functions after removal of the antibiotic. *BBA - Gene Struct. Expr.* **1009**, 39–46 (1989).
149. Ojima, I. New cyclization reactions in organic syntheses. *Pure Appl. Chem.* **74**, 159–166 (2002).
150. 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 (2010).
151. Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F. & Leadlay, P. F. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **25**, 447–453 (2007).
152. Oostu, K., Matsumoto, T., Harada, S. & Kishi, T. Antitumor and immunosuppressive activities of lankacidin-group antibiotics: structure-activity relationships. *Cancer Chemother. Rep.* **59**, 919–928 (1975).
153. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

154. Panjkovich, A. & Svergun, D. I. Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys. Chem. Chem. Phys.* **18**, 5707–5719 (2016).
155. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
156. Pappenberger, G., Benz, J., Gsell, B., Hennig, M., Ruf, A., Stihle, M., Thoma, R. & Rudolph, M. G. Structure of the Human Fatty Acid Synthase KS-MAT Didomain as a Framework for Inhibitor Design. *J. Mol. Biol.* **397**, 508–519 (2010).
157. Patterson, A. L. A Direct Method for the Determination of the Components of Interatomic Distances in Crystals. *Zeitschrift für Krist. - Cryst. Mater.* **90**, 517–542 (1935).
158. Paul, V. J., Frautschy, S., Fenical, W. & Neelson, K. H. Antibiotics in microbial ecology: Isolation and structure assignment of several new antibacterial compounds from the insect-symbiotic bacteria *Xenorhabdus* spp. *J. Chem. Ecol.* **7**, 589–597 (1981).
159. Peltier, K. The new macrolactones: clarithromycin (Biaxin) and azithromycin (Zithromax). *Conn. Med.* **56**, 371–374 (1992).
160. Pérez, J. & Koutsioubas, A. Memprot: A program to model the detergent corona around a membrane protein based on SEC-SAXS data. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 86–93 (2015).
161. Petoukhov, M. V. & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–50 (2005).
162. 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).
163. Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996–1047 (2010).

164. Plestil, J., Pospisil, H., Ostanevich, Y. M. & Degovics, G. Molecular-weight determination from small-angle scattering without absolute intensities. Advantages and limitations. *J. Appl. Crystallogr.* **24**, 659–664 (1991).
165. Preisig-Müller, R., Gehlert, R., Melchior, F., Stietz, U. & Kindl, H. Plant polyketide synthases leading to stilbenoids have a domain catalyzing malonyl-CoA:CO₂ exchange, malonyl-CoA decarboxylation, and covalent enzyme modification and a site for chain lengthening. *Biochemistry* **36**, 8349–8358 (1997).
166. 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).
167. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. a. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
168. Qin, X., Sun, L., Wen, X., Yang, X., Tan, Y., Jin, H., Cao, Q., Zhou, W., Xi, Z. & Shen, Y. Structural insight into unique properties of protoporphyrinogen oxidase from *Bacillus subtilis*. *J. Struct. Biol.* **170**, 76–82 (2010).
169. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–95 (1998).
170. Rahman, A. S., Hothersall, J., Crosby, J., Simpson, T. J. & Thomas, C. M. Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series. *J. Biol. Chem.* **280**, 6399–408 (2005).
171. Rambo, R. P. & Tainer, J. A. Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology SAXS: small-angle X-ray scattering. *Annu. Rev. Biophys* **42**, 415–41 (2013).
172. Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2008).

173. Rix, U., Fischer, C., Remsing, L. L. & Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **19**, 542–580 (2002).
174. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 (2014).
175. Robinson, R. A Synthesis of Tropinone. *J. Chem. Soc.* **111**, 762–768 (1917).
176. Rossmann, M. G. The molecular replacement method. *Acta Crystallogr. Sect. A* **46**, 73–82 (1990).
177. Rossmann, M. G. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24–31 (1962).
178. 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).
179. Roze, L. V., Chanda, A. & Linz, J. E. Compartmentalization and molecular traffic in secondary metabolism: A new understanding of established cellular processes. *Fungal Genet. Biol.* **48**, 35–48 (2011).
180. Rudd, B. A.; Hopwood, D. A. Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* **114**, 35–43 (1979).
181. Scaglione, J. B., Akey, D. L., Sullivan, R., Kittendorf, J. D., Rath, C. M., Kim, E. S., Smith, J. L. & Sherman, D. H. Biochemical and structural characterization of the tautomycetin thioesterase: Analysis of a stereoselective polyketide hydrolase. *Angew. Chemie - Int. Ed.* **49**, 5726–5730 (2010).
182. Schell, U., Haydock, S. F., Kaja, A. L., Carletti, I., Lill, R. E., Read, E., Sheehan, L. S., Low, L., Fernandez, M.-J., Grolle, F., McArthur, H. a I., Sheridan, R. M., Leadlay, P. F., Wilkinson, B. & Gaisser, S. Engineered biosynthesis of hybrid macrolide polyketides containing D-angolosamine and D-mycaminose moieties. *Org. Biomol. Chem.* **6**, 3315–27 (2008).

183. 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).
184. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* **99**, 14083–14088 (2002).
185. 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).
186. Serre, L., Verbree, E. C., Dauter, Z., Stuitje, a R. & Derewenda, Z. S. The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 1.5-Å resolution. Crystal structure of a fatty acid synthase component. *The Journal of biological chemistry* **270**, 12961–12964 (1995).
187. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479–485 (2010).
188. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).
189. Sieber, S. a., Linne, U., Hillson, N. J., Roche, E., Walsh, C. T. & Marahiel, M. a. Evidence for a monomeric structure of nonribosomal peptide synthetases. *Chem. Biol.* **9**, 997–1008 (2002).
190. Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, 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**, 1145–1153 (2005).
191. Smith, S., Witkowski, A. & Joshi, A. K. Structural and functional organization of the animal fatty acid synthase. *Prog. Lipid Res.* **42**, 289–317 (2003).

192. Squire, C. M., Goss, R. J. M., Hong, H., Leadlay, P. F. & Staunton, J. Catalytically active tetramodular 6-deoxyerythronolide B synthase fusion proteins. *Chembiochem* **4**, 1225–1228 (2003).
193. Stachelhaus, T. & Walsh, C. T. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry* **39**, 5775–5787 (2000).
194. Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J. & Hranueli, D. ClustScan: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.* **36**, 6882–6892 (2008).
195. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
196. 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**, 1345–9 (2004).
197. Straight, P. D., Fischbach, M. a, Walsh, C. T., Rudner, D. Z. & Kolter, R. A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 305–310 (2007).
198. Sugawara, A., Sueki, A., Hirose, T., Nagai, K., Gouda, H., Hirono, S., Shima, H., Akagawa, K. S., Omura, S. & Sunazuka, T. Novel 12-membered non-antibiotic macrolides from erythromycin A; EM900 series as novel leads for anti-inflammatory and/or immunomodulatory agents. *Bioorganic Med. Chem. Lett.* **21**, 3373–3376 (2011).
199. Suhre, K. & Sanejouand, Y. H. Elnémo: A normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32**, 610–614 (2004).

200. Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H., Nimi, O. & Kinashi, H. Identification of two polyketide synthase gene clusters on the linear plasmid pSLA2-L in *Streptomyces rochei*. *Gene* **246**, 123–131 (2000).
201. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates D. Svergun, C. Barberato and M. H. J. Koch. 768–773 (1995).
202. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
203. Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. **76**, 2879–2886 (1999).
204. Svergun, D. I., Petoukhov, M. V & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
205. Tang, G.-L., Cheng, Y.-Q. & Shen, B. Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem. Biol.* **11**, 33–45 (2004).
206. 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).
207. 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).
208. 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**, 11124–9 (2006).
209. 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**, 11124–9 (2006).
210. Tatsuno, S., Arakawa, K. & Kinashi, H. Analysis of Modular-iterative Mixed Biosynthesis of Lankacidin by Heterologous Expression and Gene Fusion. *J. Antibiot. (Tokyo)*. **60**, 700–708 (2007).

211. Tatsuno, S., Arakawa, K. & Kinashi, H. Extensive Mutational Analysis of Modular-Iterative Mixed Polyketide Biosynthesis of Lankacidin in *Streptomyces rochei*. *Biosci. Biotechnol. Biochem.* **73**, 2712–2719 (2009).
212. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 325–338 (2010).
213. Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A. W., Smith, C. P. & Micklefield, J. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew. Chemie - Int. Ed.* **51**, 7181–7184 (2012).
214. Thomas, I., Martin, C. J., Wilkinson, C. J., Staunton, J. & Leadlay, P. F. Skipping in a hybrid polyketide synthase: Evidence for ACP-to-ACP chain transfer. *Chem. Biol.* **9**, 781–787 (2002).
215. Till, M. & Race, P. R. Progress challenges and opportunities for the re-engineering of trans-AT polyketide synthases. *Biotechnol. Lett.* **36**, 877–888 (2014).
216. Till, M. & Race, P. R. The Assembly Line Enzymology of Polyketide Biosynthesis. *Methods Mol. Biol.* **1401**, 31–49 (2016).
217. Traitcheva, N., Jenke-Kodama, H., He, J., Dittmann, E. & Hertweck, C. Non-colinear polyketide biosynthesis in the aureothin and neo-aureothin pathways: An evolutionary perspective. *ChemBioChem* **8**, 1841–1849 (2007).
218. Tran, L., Tosin, M., Spencer, J. B., Leadlay, P. F. & Weissman, K. J. Covalent linkage mediates communication between ACP and TE domains in modular polyketide synthases. *ChemBioChem* **9**, 905–915 (2008).
219. Tropf, S., Karcher, B., Schroder, G. & Schroder, J. Reaction mechanisms of homodimeric plant polyketide synthases (stilbene and chalcone synthase). A single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones, and 6'-deoxychalcones. *Journal of Biological Chemistry* **270**, 7922–7928 (1995).

220. 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).
221. Tsai, S. C. & Ames, B. D. Chapter 2 Structural Enzymology of Polyketide Synthases. *Methods Enzymol.* **459**, 17–47 (2009).
222. Tsai, S. C., Lu, H., Cane, D. E., Khosla, C. & Stroud, R. M. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41**, 12598–12606 (2002).
223. 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).
224. Tuan, J. S., Weber, J. M., Stayer, M. J., Lenng, J. O., Donadio, S. & Katz, L. Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene* **90**, 21–29 (1990).
225. Del Vecchio, F., Petkovic, H., Kendrew, S. G., Low, L., Wilkinson, B., Lill, R., Cortés, J., Rudd, B. a M., Staunton, J. & Leadlay, P. F. Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **30**, 489–494 (2003).
226. Vidgren, J., Svensson, L. a & Liljas, a. Crystal structure of catechol O-methyltransferase. *Nature* **368**, 354–358 (1994).
227. Volkov, V. V & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
228. Wagner, D. T., Stevens, D. C., Mehaffey, M. R., Manion, H. R., Taylor, R. E., Brodbelt, J. S. & Keatinge-Clay, A. T. α -Methylation Follows Condensation in the Gephyronic Acid Modular Polyketide Synthase. *Chem. Commun.* (2016).

229. Wahab, H. a, Yam, W. K., Samian, M.-R. & Najimudin, N. Refinement of a low-resolution crystal structure to better understand erythromycin interactions on large ribosomal subunit. *J. Biomol. Struct. Dyn.* **26**, 131–146 (2008).
230. Walsh, C. T. & Wencewicz, T. a. Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* **30**, 175–200 (2012).
231. Wang, B. C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112 (1985).
232. Weissman, K. J. Introduction to polyketide biosynthesis. *Methods Enzymol.* **459**, 3–16 (2009).
233. Weissman, K. J. Natural Product Reports Genetic engineering of modular PKSs : from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33**, 203–230 (2016).
234. Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.* **32**, 436–53 (2015).
235. 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**, 13849–13855 (1997).
236. Weissman, K. J., Hong, H., Popovic, B. & Meersman, F. Evidence for a Protein-Protein Interaction Motif on an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Chem. Biol.* **13**, 625–636 (2006).
237. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3**, 925–936 (2005).
238. 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**, 1340–51

239. 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. & Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).
240. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
241. 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).
242. Witkowski, A., Joshi, A. K., Lindqvist, Y. & Smith, S. Conversion of a ??-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine. *Biochemistry* **38**, 11643–11650 (1999).
243. Wlodawer, a & Hodgson, K. O. Crystallization and crystal data of monellin. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 398–399 (1975).
244. Wong, F. T., Jin, X., Mathews II, Cane, D. E. & Khosla, C. Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* **50**, 6539–6548 (2011).
245. Worsham, L. M. S., Earls, L., Jolly, C., Langston, K. G., Trent, M. S. & Ernst-Fonberg, M. Lou. Amino acid residues of Escherichia coli acyl carrier protein involved in heterologous protein interactions. *Biochemistry* **42**, 167–176 (2003).
246. 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).

247. Xie, X., Garg, A., Keatinge-Clay, A. T., Khosla, C. & Cane, D. E. Epimerase and Reductase Activities of Polyketide Synthase Ketoreductase Domains Utilize the Same Conserved Tyrosine and Serine Residues. *Biochemistry* **55**, 1179–1186 (2016).
248. Xu, J., Wan, E., Kim, C. J., Floss, H. G. & Mahmud, T. Identification of tailoring genes involved in the modification of the polyketide backbone of rifamycin B by *Amycolatopsis mediterranei* S699. *Microbiology* **151**, 2515–2528 (2005).
249. Yadav, G., Gokhale, R. S. & Mohanty, D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335–363 (2003).
250. Yadav, G., Gokhale, R. S. & Mohanty, D. Towards Prediction of Metabolic Products of Polyketide Synthases: An In Silico Analysis. *PLoS Comput. Biol.* **5**, (2009).
251. Yamamoto, S., He, Y., Arakawa, K. & Kinashi, H. γ -butyrolactone-dependent expression of the *Streptomyces* antibiotic regulatory protein gene *srrY* plays a central role in the regulatory cascade leading to lankacidin and lankamycin production in *Streptomyces rochei*. *J. Bacteriol.* **190**, 1308–1316 (2008).
252. Yatsukawa, Y., Fujita, K., Nakamura, M., Watai, M., Murayama, M. & Maitani, T. Simultaneous Determination of Sedecamycin and Terdecamycin in Livestock Products by LC/MS. *J. Food Hyg. Soc. Japan* **49**, 239–243 (2008).
253. Yeh, E., Kohli, R. M., Bruner, S. D. & Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBioChem* **5**, 1290–1293 (2004).
254. Zeng, J., Wagner, D. T., Zhang, Z., Moretto, L., Addison, J. D. & Keatinge-Clay, A. T. Portability and Structure of the Four-Helix Bundle Docking Domains of *trans*-Acyltransferase Modular Polyketide Synthases. *ACS Chem. Biol.* [acschembio.6b00345](https://doi.org/10.1021/acscchembio.6b00345) (2016). doi:10.1021/acscchembio.6b00345
255. Zhang, K. Y. J., Cowtan, K. & Main, P. Combining Constraints for Electron-Density Modification. *Methods Enzymol.* **277**, 53–64 (1997).

256. Zhang, L., Liu, W., Xiao, J., Hu, T., Chen, J., Chen, K., Jiang, H. & Shen, X. Malonyl-CoA: acyl carrier protein transacylase from *Helicobacter pylori*: Crystal structure and its interaction with acyl carrier protein. *Protein Sci* **16**, 1184–1192 (2007).
257. 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).
258. 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).

Les polykétides sont des molécules complexes possédant des rôles thérapeutiques divers (antibiotiques, anticancéreux, immunosuppresseurs, etc.). Ces composés sont synthétisés par les polykétides synthases (PKS) qui présentent donc un intérêt certain. Une stratégie prometteuse et en développement, appelée biologie synthétique, consiste en l'ingénierie protéique des PKS pour produire de nouveaux polykétides. Cependant, un prérequis au succès de cette méthode est la compréhension du fonctionnement des PKS. En effet, les PKS sont des systèmes complexes composés de plusieurs sous-unités. Celles-ci comportent chacune un ou plusieurs modules responsable d'un cycle d'extension du polykétide. Les modules sont composés également de plusieurs domaines assurant chacun un rôle biosynthétique.

Il est ainsi nécessaire de comprendre comment s'opère la communication entre domaines au sein d'un module. Pour cela, nous avons étudié le module 5 de la PKS synthétisant la virginiamycine M par une approche combinant la diffusion des rayons X aux petits angles (SAXS), la résonance magnétique nucléaire (RMN) et la modélisation par homologie. Ainsi nous sommes parvenus à caractériser la structure en solution du module 5, mais également à positionner les structures à haute résolution des domaines à l'intérieur de l'enveloppe SAXS. De plus, notre étude de la dynamique du module 5 a montré que les deux domaines acyl carrier protein (ACP) composant le module semblent non-équivalents, et que l'ACP_{5b} doit être doté d'une certaine mobilité afin d'être fonctionnel, ceci dû à la flexibilité du linker reliant les deux ACP.

L'interaction entre les sous-unités consécutives est également primordiale pour assurer la fidélité de la synthèse des polykétides. Ces interactions sont assurées, au moins en partie, par des domaines de petite taille appelés « domaine de docking (DD) ». Jusqu'alors, les DD avaient été caractérisés uniquement chez les PKS de type cis-AT. Nous avons caractérisé une nouvelle classe de DD, la première chez les PKS trans-AT. Grâce à une approche pluridisciplinaire, nous avons montré que l'un des DD constitue probablement une protéine intrinsèquement désordonnée (IDP) : son interaction avec le DD partenaire induirait son repliement. Nous avons résolu la structure RMN d'une protéine de fusion entre les deux DD, affichant une nouvelle topologie pour une interaction protéine-protéine. Cette interface de docking a ensuite été replacée dans son contexte modulaire par SAXS. Contrairement aux autres DD qui ne forment qu'une seule interface de docking, ces DD forment deux complexes de docking à l'intersection des deux sous-unités. Nos données SAXS nous avons également permis de proposer un modèle de l'interface créée entre deux sous-unités PKS dans laquelle une chambre réactionnelle est formée, qui pourrait servir à protéger des intermédiaires réactifs de polykétide.

Des enzymes post-PKS interviennent suite à la synthèse du polykétide pour maturer ce dernier. Cette étape est d'une importance considérable puisqu'elle confère la structure et l'activité finale du polykétide. Dans ce contexte, nous avons mené une étude structure/fonction de l'enzyme post-PKS catalysant la macrocyclisation de l'antibiotique lankacidine C. Après un phasage par SAD sur la protéine sélénée, nous avons résolu la structure de l'enzyme en complexe avec des analogues de substrats, puis procéder à la conception de mutants pour résoudre la structure de l'enzyme avec son substrat naturel.

Mots clés : polykétide, PKS, domaine de docking, SAXS, enzyme post-PKS, cristallographie des protéines

Polyketides are structurally complex molecules which exhibit diverse therapeutic activities (antibiotic, antitumor, immunosuppressant...). These compounds and the enzymes responsible for their synthesis, the polyketide synthases (PKSs), are thus of significant biomedical interest. An emerging though promising strategy for the generation of novel polyketides is the engineering of the PKS proteins, an approach called synthetic biology. Nevertheless, a fundamental understanding of the mode of operation of the PKS enzymes is directly correlated with the success of this methodology. Indeed, PKSs are molecular-scale assembly lines which are composed of several subunit, each of which includes one or several modules catalyzing a polyketide elongation cycle. The module themselves are composed of several domains each with a specific role in the biosynthesis.

It is therefore necessary to understand how the domains within a module communicate with each other. To address this question, we studied module 5 of the PKS responsible for virginiamycin M using an approach combining small angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) and homology modeling. This strategy allowed us to characterize the solution structure of module 5, but additionally to position high-resolution domain structures inside the SAXS envelopes. We also studied the dynamics of module 5, demonstrating that the two component acyl carrier proteins (ACPs) appear to be non-equivalent, and that the function of ACP_{5b} is likely dependent on the mobility conferred on it by the flexible linker between the two domains.

The interaction between the consecutive subunits is also critical to ensuring the fidelity of polyketide synthesis. This communication is assured, at least in part, by small domains called docking domains (DD), which have to date only been characterized from cis-AT PKS. Here, we have identified a new class of DD, the first shown to be present in trans-AT PKSs. Using a multidisciplinary approach, we have demonstrated that the N-terminal DD (VirFG^NDD) is likely to be an intrinsically disordered protein (IDP), whose interaction with its partner DD (VirA^CDD) induces its folding. We have solved the NMR structure of a fusion protein between the two DD, revealing a new topology for a protein-protein interaction. This docking interface was then placed into its modular context by SAXS. In contrast to the other classes of DD which form a single docking complex, this new type of DD gives rise to two docking interfaces at the intersubunit junction. Finally, our SAXS data have allowed us to propose a model for the complete interface between two PKS subunits in which a reaction chamber is formed, which may allow reactive polyketide intermediates to be protected.

Post-PKS enzymes catalyze maturation of the polyketide after its release from the last module of the PKS. This processing is critical as it yields the final structure and activity of the polyketide. In this context, we conducted a structure/function study of the post-PKS enzyme catalyzing the macrocyclisation of the antibiotic lankacidine C. After phasing by SAD using a seleniated protein, we solved the structure of the enzyme in complex with substrate analogues. We then proceeded to site-directed mutagenesis of specific residues, in order to solve the structure of the enzyme in complex with its natural substrate.

Key words: polyketide, PKS, docking domains, SAXS, post-PKS enzyme, protein crystallography